



UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE CIENCIAS EXACTAS

DEPARTAMENTO DE CIENCIAS BIOLÓGICAS

Laboratorio de Investigación de Proteínas Vegetales (LIPROVE)

Cátedra de Química Medicinal

***Peptidasas vegetales como biocatalizadores
para la obtención de intermediarios de
síntesis de compuestos con potencial actividad
farmacológica***

María Eugenia Tavarone

Directora: Dra. SUSANA R. MORCELLE DEL VALLE

Codirector: Dr. ALAN TALEVI

Laboratorio de Procesos Biotecnológicos – Año 2011

1. Introducción

Marco histórico

Desde el comienzo de los tiempos el hombre se ha empeñado en encontrar cura y tratamiento a sus males, y de esta manera comenzó a volcarse a la exploración de vegetales, minerales y animales como fuente de soluciones, al principio mágicas y pseudo-empíricas, que gradualmente se volvieron más racionales, ligándose así al pensamiento científico y la experimentación. Los primeros indicios serios de la búsqueda y modificación de un compuesto con actividad biológica asociada, con potencial de corregir, tratar o prevenir patologías, es decir, con actividad farmacológica, fue el aislamiento de la salicina (siglo XIX). Ésta luego fue modificada químicamente para su introducción al mercado como ácido acetil salicílico (aspirina) en 1899 por la Farbenfabriken in Bayer, de Alemania.

Actualmente la Farmacología se encuentra en la era de la Farmacología Inversa, en la que se identifican los posibles genes implicados en las patologías, se expresan receptores o enzimas generalmente en diversas especies transgénicas (ratones knock-out, microorganismos como *Escherichia coli.*, etc.) y luego se estudian los posibles blancos del fármaco y sus posibles puntos de influencia química. Hoy se plantea la posibilidad de diseñar un medicamento utilizando un mecanismo de aproximación al proceso patológico, es decir, conociendo la base biológica de una enfermedad o de un desarreglo metabólico. Cuando se conoce este proceso en su base molecular y se pueden definir las moléculas implicadas en el mismo, es posible diseñar medicamentos que interactúen con la molécula responsable, de tal forma que la modifique y se modifique asimismo la patología.

El descubrimiento de un nuevo medicamento y su desarrollo posterior son dos fases que condicionan el logro de un nuevo producto, útil en la terapéutica. Todo el proceso, desde su obtención hasta su registro comprende un total de aproximadamente 10 años de investigación, con un costo promedio de 800 millones de dólares por cada nuevo medicamento que sale al mercado (Miller P, 2005). Posteriormente, las grandes industrias farmacéuticas recuperan con creces la inversión inicial, con el consecuente impacto en el

valor de los medicamentos, transformándolos en un negocio millonario y despojándolos del significado social que debieran tener.

En este mismo sentido, el diseño racional de fármacos constituye una herramienta indispensable en el desarrollo actual de nuevos medicamentos, contribuyendo a un aumento de las posibilidades de éxito y a una disminución de los costos relacionados con el proceso de generación de los mismos. Por otro lado, el empleo de recursos naturales autóctonos como fuente de moléculas intermediarias en la síntesis de compuestos de interés farmacológico, o como material de partida para la obtención de catalizadores naturales, enzimas capaces de sintetizar dichos compuestos, es fundamental, no solo por permitir la reducción de los costos, sino también por permitir la independencia de problemas socio-políticos y económicos que podrían limitar y condicionar el desarrollo de la industria farmacéutica nacional. Además, la diversidad en la flora de nuestro país nos coloca en una situación de partida ventajosa en este aspecto.

El presente trabajo aspira, en forma modesta y limitada, a plasmar lo expresado anteriormente en el estudio de derivados de N^α -benzoil-L-arginina como posibles compuestos con actividad inhibitoria sobre receptores de tipo integrina, mediante el empleo de la metodología QSAR (Quantative Structure-Activity Relationships), y su obtención, empleando peptidasas vegetales de origen autóctono.

Marco biológico: las integrinas

La mayor parte de las funciones fisiológicas son reguladas por interacciones entre péptidos o proteínas y receptores u otras proteínas. Muy a menudo, tales ligandos ejercen su actividad biológica a través de regiones relativamente pequeñas y plegadas de su superficie, y por lo tanto sus acciones pueden ser reproducidas por drogas mucho más pequeñas pero que conservan estas superficies bioactivas localizadas (Gentilucci *et al.*, 2006). Un ejemplo típico son las integrinas (Giancotti *et al.*, 1999) y sus ligandos naturales, la matriz extracelular, fibronectina, vitronectina, fibrinógeno, laminina, el factor de von Willebrand, etc.

Las integrinas son una superfamilia de receptores de membrana, glicoproteicos y heterodiméricos, expresada abundantemente en la superficie celular. Son proteínas de adhesión que participan en eventos de adhesión célula-célula y célula-matriz que tienen un rol importante durante el desarrollo, la patogénesis y durante la migración y la señalización celular (Pytela *et al.*, 1986; Ruoslahti *et al.*, 1986; Ruoslahti *et al.*, 1987).

Las integrinas son expresadas en una amplia variedad de tipos celulares como monocitos, osteoclastos, linfocitos activados, células endoteliales, plaquetas, células de músculo liso, melanomas y carcinomas de mama. La habilidad de interferir selectivamente con la función de receptores de adhesión celular a través de moléculas pequeñas ofrece muchas oportunidades para la intervención terapéutica de enfermedades tan diversas como trombosis, inflamación, osteoporosis, retinopatía diabética, restenosis y cáncer (Phillips *et al.*, 1988; Kieffer & Phillips, 1990; Phillips *et al.*, 1991; Hynes, 1992; Cox *et al.*, 1994; Mousa, 1996). Por ejemplo, las integrinas $\alpha_v\beta_3$, expresadas en varios tipos celulares (entre ellos células endoteliales, osteoclastos, y células de melanoma), reconocen sus ligandos al interactuar con un motivo de unión muy corto, tripeptídico, el motivo RGD (Arg-Gly-Asp) (Figura 1.1).

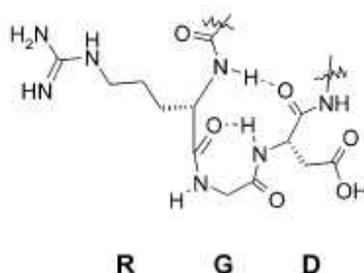


Figura 1.1: Estructura del tripeptido RGD.

El tripeptido RGD es una secuencia característica común de muchos ligandos de la integrina (Ruoslahti, 1996). Por lo tanto, una serie de antagonistas de integrinas basados en la secuencia RGD se han diseñado como agentes terapéuticos para el tratamiento de la angiogénesis (Stupack *et al.*, 2004), tumores, (Haubner *et al.*, 1997), osteoporosis, (Hartman & Duggan, 2000), etc. Sin embargo, por las similitudes estructurales dentro de la familia de las integrinas y entre sus respectivos ligandos, las integrinas α_v , así como las

integrinas $\alpha_v\beta_1$, $\alpha_{vIII}\beta_1$ y $\alpha_{IIb}\beta_3$, reconocen la misma secuencia RGD. Por consiguiente, la cuestión de la selectividad constituye una importante preocupación en el diseño de agentes terapéuticos potenciales. Por esta razón, los *RGD mimetics* (Henry *et al.*, 2002; Miller *et al.*, 2000; Cacciari & Spallato, 2005) han sido ampliamente modificados mediante la inclusión de algún tipo de estructuras para sesgar la conformación (núcleos rígidos, estructuras cíclicas, etc).

Debido a que tanto la alta biodisponibilidad como el bajo peso molecular son características necesarias para el desarrollo de drogas, la búsqueda de antagonistas efectivos se reorientó de péptidos grandes y cíclicos a compuestos pequeños y estructuralmente más simples, así como a péptidos miméticos (*peptide mimetics*) aquirales e incluso análogos no peptídicos. Esto condujo a la identificación de numerosos antagonistas selectivos de pequeño tamaño, los que llegaron a probarse en ensayos clínicos (Henry *et al.*, 2002; Cacciari. & Spallato, 2005). En consecuencia, se decidió enfocar este trabajo de investigación al diseño de pequeños y sencillos análogos dipeptídicos de RGD.

1.1. Síntesis enzimática empleando proteasas vegetales como biocatalizadores

1.1.1. Generalidades

Las enzimas proteolíticas, también conocidas como proteasas o peptidasas, son hidrolasas que catalizan la hidrólisis reversible del enlace peptídico de las proteínas. Se encuentran naturalmente en todos los organismos vivos, ya que la proteólisis juega un rol importante en muchos procesos biológicos tales como la digestión, el recambio de proteínas y la defensa frente a patógenos.

1.1.2. Interacción enzima-sustrato

En general, las proteasas se unen al sustrato apropiado a través de los denominados sitios activos que permiten luego catalizar la hidrólisis de la unión peptídica específica (Mc Donald, 1985). Según la nomenclatura introducida por Schechter y Berger (1967) y hoy utilizada ampliamente, la zona de unión del sitio activo puede subdividirse en subsitios o sitios específicos (S), cada uno de los cuales permite la ubicación de un residuo

aminoacídico del sustrato o péptido (P). Los subsitios se ubican a ambos lados del sitio activo y a partir del mismo se numeran en forma correlativa, bien respecto en dirección al N-terminal (S_1, S_2, S_3 , etc) o bien hacia el C-terminal (S'_1, S'_2, S'_3 , etc.). En cuanto al sustrato, los aminoácidos que interactúan con la proteasa son numerados en relación a los subsitios en los cuales se ubican respecto del N-terminal (P_1, P_2, P_3 , etc.) o hacia el C-terminal (P'_1, P'_2, P'_3 , etc.) (Figura 1.2).

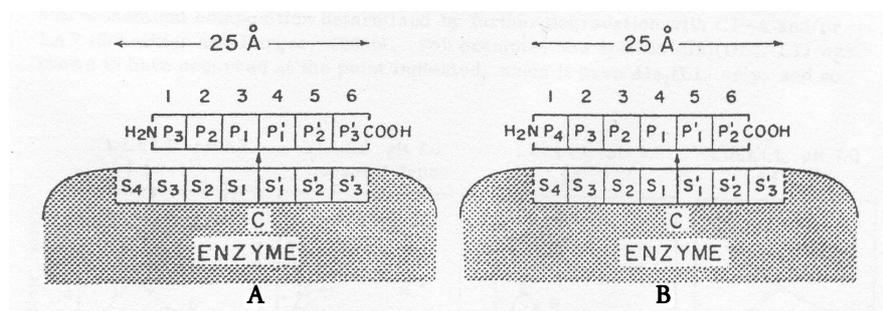


Figura 1.2: Interacción enzima-sustrato. El sitio activo de la enzima está compuesto de siete subsitios, (S_1-S_4 y $S'_1-S'_3$) localizados a ambos lados del sitio catalítico C. Las posiciones P del sustrato hexapeptídico se cuentan desde el punto de corte, teniendo así la misma numeración que los subsitios que ocupan. A. Este complejo enzima-sustrato libera como producto dos moléculas de tripéptido. B. Este segundo complejo libera una molécula de tetrapéptido y una de dipéptido. (Schechter y Berger, 1967).

1.1.3. Clasificación

El Comité de Nomenclatura de la Unión Internacional de Bioquímica y Biología Molecular (NC-IUBMB) en 1984 recomendó utilizar el término *peptidasa* para el grupo de hidrolasas de unión peptídica asignándoles el código E.C. 3.4. (Barret *et al.*, 2004). Las *peptidasas* a su vez comprenden dos grandes grupos de enzimas: las endopeptidasas, que rompen uniones peptídicas en distintos puntos del interior de la proteína, y las exopeptidasas, que remueven secuencialmente aminoácidos desde los extremos carboxilo o amino (Barrett, 2001).

Dada la dificultad de definir las en base a su especificidad por el sustrato, las proteasas se clasifican en base a su mecanismo catalítico tal como lo propusiera originalmente Hartley (1960). Se diferencian así cuatro clases: serínicas (EC 3.4.21), cisteínicas (EC 3.4.22), aspárticas (EC 3.4.23) y metalopeptidasas, de acuerdo al

aminoácido o metal divalente involucrado directamente en la catálisis. Cada una posee un mecanismo catalítico distintivo; sin embargo, es de destacarse que las proteasas serínicas y cisteínicas forman complejos covalentes entre la enzima y el sustrato, en tanto que las metaloproteasas y las aspárticas no lo hacen. Posteriormente fueron reportadas proteasas glutámicas (Kakudo *et al.*, 1992) y más recientemente las peptidasas treonínicas (Barrett *et al.*, 1998).

1.1.4. Peptidasas cisteínicas

Dentro de las peptidasas cisteínicas (EC 3.4.22) pueden mencionarse numerosas proteasas vegetales extensamente estudiadas tales como papaína, ficina, actinidina y bromelaína. La actividad catalítica de este tipo de peptidasas ocurre a través de la formación de un intermediario covalente e involucra un residuo de histidina «159» y uno de cisteína «25» (numeración de acuerdo a papaína). Este último residuo, a su vez, se encuentra involucrado en un equilibrio tautomérico entre las formas neutra y zwitteriónica (Figura 1.3). En la primera etapa (etapa de acilación) el sulfuro aniónico actúa como nucleófilo sobre el grupo carbonilo del sustrato, formándose así un intermediario covalente que se estabiliza a través de la formación de un par iónico con el grupo vecino imidazol de la histidina. La formación de este intermediario procede a través de un estado intermedio de transición tetraédrico cargado negativamente (Figura 1.3). Durante la segunda etapa o etapa de deacilación, el intermedio acil-enzima es hidrolizado por una molécula de agua (nucleófilo) para liberar el péptido y restaurar el grupo catalíticamente activo de la proteasa. La deacilación involucra también la formación de un estado de transición tetraédrico, que se produce a través de la reacción inversa a la vía de acilación. Estas enzimas son inactivadas por reactivos bloqueantes de los grupos sulfhidrilo (por conversión en puentes disulfuro) y tienen la capacidad de reactivarse en presencia de agentes reductores (Morcelle, 2004).

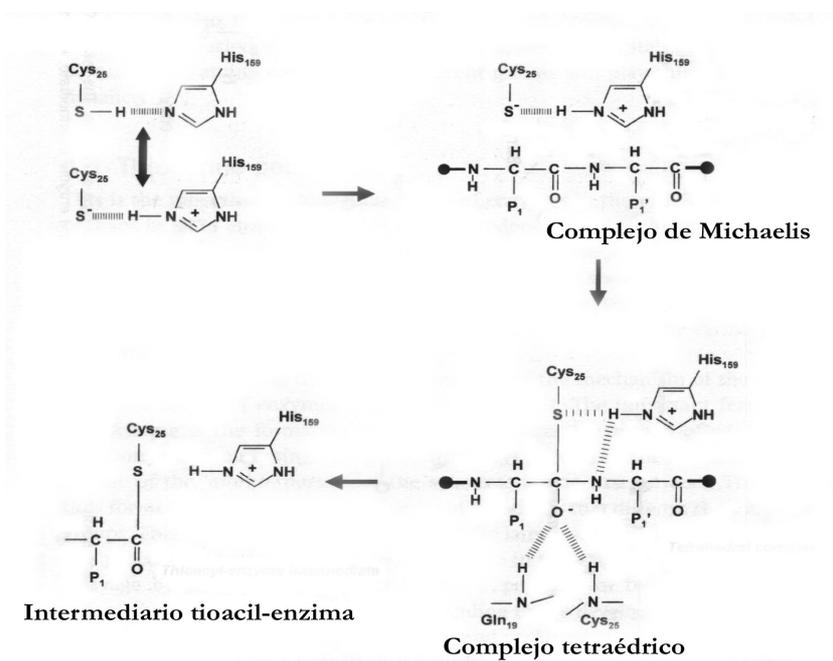


Figura 1.3: Mecanismo catalítico de las proteasas cisteínicas.

1.1.5. Rol de las peptidasas en los vegetales

Las proteasas vegetales están involucradas en casi todos los aspectos del crecimiento y el desarrollo de la planta, incluyendo germinación (donde intervienen en la degradación de proteínas de reserva), ritmos circadianos, senescencia y muerte celular programada. Desempeñan un papel clave en la regulación de los procesos biológicos, como el reconocimiento de agentes patógenos y plagas y en la eficaz inducción de respuestas de defensa, atacando a los organismos invasores. Todo esto explica el hecho de que las proteasas ocupen del 1 al 5% del genoma. Por otro lado, estas enzimas son capaces de activar a los receptores blanco de proteasas y de esta manera actuar como importantes agentes farmacológicos y toxicológicos (Domsalla & Melzig, 2008).

En general, aquellas proteasas vinculadas al desarrollo y crecimiento se encuentran en muy bajas cantidades en las plantas. Sin embargo, algunas especies vegetales poseen una gran concentración de proteasas sumamente activas en ciertos tejidos. Un ejemplo de esto lo constituye el látex obtenido a partir de plantas de diversas familias en las que el contenido de enzimas proteolíticas puede superar el 50% de las proteínas totales (Morcelle, 2004).

1.1.6. Proteasas de látex

El látex es considerado una fuente importante de proteasas vegetales y es bien conocida su utilización tanto en medicina popular como en diferentes procesos industriales. Casi la mitad de las enzimas comerciales disponibles son proteasas, utilizadas frecuentemente en la elaboración de alimentos, cervezas, quesos, tiernización de carnes, fabricación de pan y en la industria textil y del cuero. Dada la amplia gama de temperatura y pH en la que mantienen sus propiedades catalíticas, estas enzimas han recibido especial atención en la industria farmacéutica y en la biotecnología. Dentro de las propiedades medicinales de las proteasas cisteínicas de látex se han descrito su poder como antihelmínticos, su aplicación en tratamientos terapéuticos de trastornos digestivos, cicatrización de heridas, inmunomodulación y alteraciones neoplásicas (Salas *et al.*, 2008).

En la planta, las proteasas contenidas en el látex probablemente intervengan en la degradación de proteínas durante el desarrollo de los laticíferos y en la promoción de la coagulación. En algunas plantas, cuando sus hojas, tallos o frutos son injuriados, la segregación de látex permite la formación de un coágulo alrededor del área dañada en pocos minutos. Este proceso es vital para la defensa contra el ataque de parásitos y patógenos (Liggieri *et al.*, 2004).

Se ha encontrado, al menos, una enzima con actividad proteolítica en más de 110 látex de diferentes especies vegetales estudiadas. La mayoría de las proteasas de látex son de tipo serínicas y cisteínicas, por lo que el látex de frutos constituye una fuente importante de proteasas.

1.1.7. Proteasas de *Araujia hortorum* (Asclepiadaceae)

En el látex de *Araujia hortorum* Fourn se han purificado mediante cromatografía de intercambio catiónico tres endopeptidasas cisteínicas denominadas araujaína h I, h II y h III, cuyos pesos moleculares son 24,0 kDa, 23,7 kDa y 23,5 kDa, respectivamente; las tres presentaron perfil de pH sobre caseína con máximo de actividad entre pH 8,0 y 9,0. Araujaína h I y h III tienen pI mayores que 9,3, mientras que araujaína h II tiene un pI algo más bajo (8,9) (Priolo *et al.*, 2000; Obregón *et al.*, 2001).

1.1.8. Aplicaciones de peptidasas vegetales en reacciones de síntesis

Existe un interés particular en el empleo de biocatalizadores para crear nuevas vías de síntesis que disminuyan el uso de reactivos costosos, así como para obtener moléculas genéricas y/o complejas empleadas en la producción de reactivos químicos, de intermediarios farmacéuticos y agroquímicos, principios activos farmacéuticos e ingredientes alimentarios. Por otro lado, nuevas oportunidades en el terreno de la biocatálisis han aparecido debido a nuevos requerimientos industriales: la limitada vigencia de las patentes de los principios activos y la rápida elaboración de procesos para compensar los altos costos de desarrollo que deben afrontar las empresas farmacéuticas, incluyendo la promoción de soluciones ecológicas que dejen rastros mínimos en el medio ambiente.

Las proteasas tienen particular interés en química orgánica debido a su alta estereoespecificidad y regioespecificidad en la catálisis. Son capaces de actuar en condiciones suaves, ya que presentan valores de pH óptimo entre 6 y 8, son fáciles de manejar, no necesitan cofactores costosos, son bastante estables, y son relativamente simples en su arquitectura molecular en la mayoría de los casos. Como catalizadores, alteran la velocidad a la cual se alcanza el equilibrio termodinámico de una reacción, sin cambiar el equilibrio en sí. Esto quiere decir que también son capaces de catalizar la reacción inversa, aunque las constantes de equilibrio para la reacción inversa, no obstante, están en el rango de 10^{-3} a 10^{-4} l/mol. Esto quiere decir que, en condiciones fisiológicas, la posición del equilibrio de la reacción está en la dirección de la hidrólisis, en tanto que la reacción inversa parecería ser insignificante *in vivo*. Para que las peptidasas puedan actuar como biocatalizadores en la formación de enlaces peptídicos debe manipularse el equilibrio de la reacción. Dichas manipulaciones consisten en dos estrategias básicas: la síntesis bajo control termodinámico y la síntesis bajo control cinético.

1.1.8.1. Síntesis bajo control termodinámico

Representa la inversa directa de la proteólisis y es aplicable a todas las peptidasas, independientemente de su mecanismo catalítico (Figura 1.4). Para esto, debe manipularse el equilibrio hacia la síntesis: esto es, influir en el equilibrio de ionización, ya que la barrera termodinámica para revertir la hidrólisis está determinada por la energía necesaria para

transferir un protón desde el grupo reactivo del nucleófilo al carboxilato negativamente cargado del dador de acilo. La energía de la barrera termodinámica para dicha transferencia puede disminuirse en presencia de solventes orgánicos que bajen la constante dieléctrica del medio, y por ende, la acidez del dador de acilo.

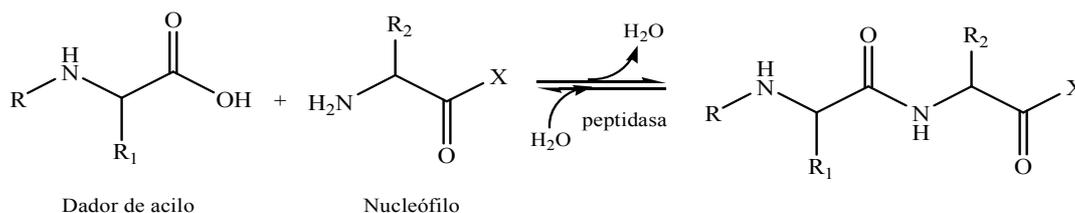


Figura 1.4: síntesis bajo control termodinámico.

Otras condiciones de reacción tales como la precipitación o extracción del producto también aumentan la eficiencia de la reacción inversa.

Como desventajas de la implementación de este tipo de estrategias deben mencionarse las bajas velocidades y el requerimiento de cantidades importantes de enzima.

1.1.8.2. Síntesis bajo control cinético

Contrariamente, la síntesis cinéticamente controlada solamente ocurre exitosamente con proteasas serínicas y cisteínicas, que forman intermediarios reactivos acil-enzima durante la catálisis. Otra característica fundamental de este tipo de síntesis es el uso de dadores de acilo levemente activados con funciones ésteres o amidas, que aceleren significativamente la velocidad de la reacción y minimicen el requerimiento de enzima (Figura 1.5).

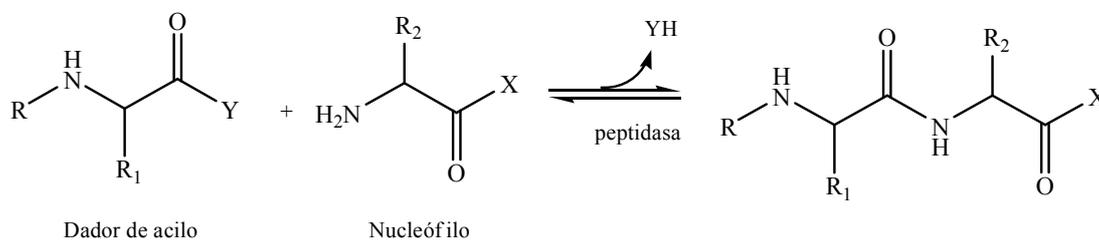


Figura 1.5: síntesis bajo control cinético

En principio, esos dadores de acilo se unen a la proteasa formando un complejo tetraédrico enzima-sustrato, llevando a un intermediario E-Ac. Los grupos carboxilato activados son esenciales para esta aplicación. El valor del método cinético para síntesis orgánica está basado en el hecho de que el intermediario Ac-E puede ser interceptado no solamente por agua, sino también por nucleófilos alternativos (NH) tales como aminas, alcoholes y tioles, los que compiten con el agua para la deacilación (Figura 1.6).

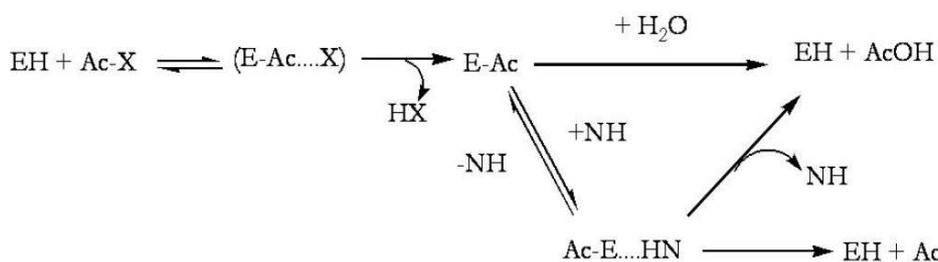


Figura 1.6: Modelo de síntesis bajo control cinético catalizada por peptidasas cisteínicas y serínicas

En este contexto, control cinético quiere decir que el producto aparece con mayor velocidad a la que desaparece y por lo tanto debe acumularse. Sin embargo, la aparición del producto llega a un máximo antes de que su propia hidrólisis (más lenta) comience a ser más importante. Entonces, si la reacción no es detenida previamente, el producto será hidrolizado cuando el dador de acilo se haya consumido, estableciéndose así el verdadero equilibrio (véase Figura 1.7).

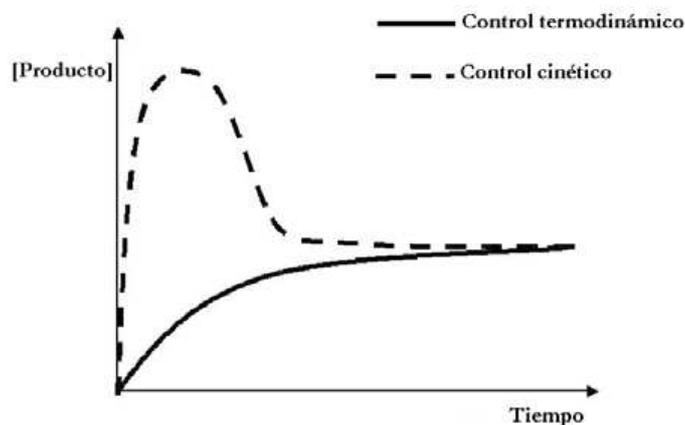


Figura 1.7: Dependencia con el tiempo en síntesis peptídica por control cinético (línea discontinua) y por control termodinámico (línea continua).

Ambos tipos de proteasas (cisteínicas y serínicas) son capaces de activar un grupo ácido carboxílico esterificado genérico o grupo amida, y transferirlo a un amplio número de nucleófilos. El éxito de las síntesis depende de muchos factores. Los parámetros fundamentales de la reacción, tales como temperatura, pH, fuerza iónica y concentración de reactivos juegan un papel importante. Bajo control cinético se puede considerar que un aumento en la concentración del nucleófilo (NH) y el pH también aumentan el rendimiento en producto. La razón es que ambas manipulaciones aumentan la eficiencia del ataque nucleofílico en el intermediario acil-enzima por el aceptor de acilo (NH) en tanto que el último resulta del hecho de que solo la forma básica libre del nucleófilo puede reaccionar con la enzima acilada. Así, el pH de la mezcla de la reacción debería ser preferiblemente más alto que el pK del nucleófilo. Por último, la enzima por sí misma afecta significativamente la eficiencia de la síntesis. En tanto que la velocidad de reacción está determinada principalmente por la especificidad de la enzima hacia el dador de acilo, es crucial la unión específica del nucleófilo al subsitio S' de la proteasa para obtener altos rendimientos. Ya que la especificidad de unión del dominio al donador de acilo y al aceptor son parámetros individuales de cada enzima, la eficiencia de síntesis y la utilidad sintética de las proteasas para la síntesis orgánica difiere de una enzima a otra (Bordusa, 2002).

1.1.9. Optimización de las reacciones catalizadas por peptidasas

Los inconvenientes principales para el empleo de las proteasas como catalizadores en síntesis son los siguientes:

- ~ formación de al menos dos productos de síntesis: el deseado y el dador de acilo hidrolizado o producto de hidrólisis (el que se obtiene por deacilación competitiva de la enzima acilada por el agua);
- ~ las especificidades y selectividades de las proteasas disponibles limitan su aplicación sintética: solo tendrán relevancia práctica las reacciones con compuestos muy relacionados con los aminoácidos por los que demuestren mayor preferencia, en tanto que aminoácidos no proteicos no serán usualmente sustratos aceptables para la estas enzimas;
- ~ riesgo de proteólisis colateral del producto;

~ la actividad y estabilidad enzimática pueden verse fuertemente afectadas por solventes, aditivos y las condiciones de reacción (pH y temperatura, por ejemplo).

Para salvar estos problemas, se han desarrollado diversas estrategias, mencionadas a continuación.

1.1.9.1. Ingeniería de medios

Se trata de la manipulación racional del medio de reacción para forzar la enzima a que catalice la reacción de síntesis (Guzmán *et al.*, 2007). El objetivo es influenciar las propiedades de la enzima alterando la naturaleza del solvente en donde la reacción se lleva a cabo (Wescott & Klibanov, 1994). En general, se trata de reemplazar al agua por solventes orgánicos, aunque se incluye en este caso las reacciones en medios acuosos congelados, fluidos supercríticos y sistemas de reacción sólidos (Bordusa, 2002). Las razones para utilizar estos medios son:

- ~ la alta solubilidad relativa de muchos compuestos en solventes orgánicos;
- ~ la habilidad de revertir reacciones hidrolíticas, convirtiéndolas a sintéticas en medios no acuosos;
- ~ la posibilidad de disminuir reacciones secundarias indeseables en medios orgánicos, tales como la hidrólisis enzimática del producto, así como la inhibición por sustrato y/o producto;
- ~ la relativa facilidad en recobrar el producto y el biocatalizador de sistemas que contengan una fase orgánica.

1.1.9.2. Ingeniería de enzimas

La ingeniería de enzimas se define como todas las estrategias destinadas a obtener un biocatalizador adaptado para llevar a cabo reacciones bajo las condiciones de síntesis, incluyendo modificaciones químicas y genéticas (Guzmán *et al.*, 2007). Así se pueden tener enzimas más estables en los medios de síntesis, disminuir la hidrólisis competitiva, la proteólisis del producto, alterar la especificidad o enantioselectividad naturales de las enzimas, etc. (Bordusa, 2002).

1.1.9.2.1. Inmovilización

La inmovilización de enzimas es un proceso en el que se confina o localiza a la enzima en una región definida del espacio, para dar lugar a formas insolubles en el medio de reacción que retienen su actividad catalítica.

Preparaciones inmovilizadas de peptidasas han sido ampliamente empleadas en síntesis (Sears & Clark, 1993; Fité *et al.*, 1998; Clapés *et al.*, 1999; Morcelle *et al.*, 2006), debido a que la inmovilización ofrece varias ventajas. No solo facilita la reutilización del biocatalizador, sino también la separación de la enzima y los productos del medio de reacción y permite el desarrollo de procesos continuos, con el consecuente incremento en la productividad y rentabilidad económica del proceso biotecnológico.

En la Tabla 1.1 se resumen las ventajas y desventajas del empleo de enzimas inmovilizadas y libres.

Tabla 1.1: Ventajas y desventajas del empleo de enzima inmovilizada. Comparación con el uso de enzima libre

Enzima libre		Enzima inmovilizada	
Ventajas	Desventajas	Ventajas	Desventajas
<ul style="list-style-type: none"> • Menor costo • Utilizable en caso de sustrato con alto peso molecular • No existen pérdidas de actividad (mayor actividad específica) 	<ul style="list-style-type: none"> • Presencia de gran cantidad de enzima remanente en el producto luego de finalizada la reacción • No es posible la reutilización de la enzima • Dificultad de un control preciso • Es necesario detener la reacción enzimática, lo cual puede afectar al producto 	<ul style="list-style-type: none"> • La enzima es reutilizable • La reacción puede terminarse separando la enzima del sustrato • El control es más preciso • Mayor estabilidad a condiciones de temperatura y pH • Puede usarse en sistemas tipo batch o continuos • Existe gran flexibilidad en el diseño de reactores 	<ul style="list-style-type: none"> • Pérdida de capacidad catalítica durante el proceso de inmovilización • Presencia de restricciones difusionales o estéricas • Inactivación con una operación continua • Mayor costo, debido al soporte y al proceso de inmovilización

Una amplia variedad de métodos de inmovilización son utilizados en la actualidad. Entre ellos se encuentra la inmovilización por adsorción, que fue utilizado en este trabajo. Este tipo de inmovilización implica la unión de la enzima al soporte por medio de interacciones iónicas, fuerzas de Van de Waals, puentes de hidrógeno, interacciones hidrofóbicas y otras. Las ventajas asociadas con este tipo de procesos son su bajo costo, fácil preparación, ausencia de cambios de especificidad enzimática y la obtención de derivados estables en medios con bajo contenido acuoso. Debido a que la unión de la enzima al soporte es débil, la obtención de derivados poco estables mecánicamente y la complejidad de optimizar las variables que controlan dicho proceso son claros inconvenientes que limitan su aplicación. La inmovilización por adsorción sobre soportes sólidos tales como poliamida o celite es un método sencillo y económico (Ferreira *et al.*, 2003). Dicho método ha permitido la aplicación eficaz de proteasas en la síntesis enzimática de péptidos en medios con bajo contenido acuoso (Xing *et al.*, 2000; Morcelle *et al.*, 2006), en la síntesis de compuestos tensioactivos (Morcelle *et al.*, 2009) y en la industria láctea.

1.1.9.2.2. Modificaciones químicas

Las modificaciones químicas de la estructura proteica son métodos económicos, fáciles de llevar a gran escala, y permiten la incorporación de aminoácidos no codificados. Entre las desventajas podemos mencionar que las aplicaciones tradicionales sufren de la falta de quimio- y regioselectividad lo que puede llevar a preparaciones enzimáticas heterogéneas e irreproducibles.

Las modificaciones más comunes son la metilación de los residuos de His (generalmente dirigidos a metilar la His catalítica) y la transformación de la Ser del sitio activo en una Cys para dar una tiol-peptidasa (como la tiol-subtilisina). Por otro lado, la unión covalente de hidratos de carbono a la superficie de las enzimas demostró mejorar su estabilidad en solventes orgánicos, y la modificación de los subsitios S2, S1 y S1' de subtilisina mediante el enlace de ligandos hidrofóbicos o cisteínas artificiales fueron exitosas (Bordusa, 2002).

1.1.9.2.3. Modificaciones genéticas

La ingeniería de proteínas mediante la manipulación de los genes que codifican la síntesis de las peptidasas consiste en hacer cambios de un sitio específico de una enzima blanco con la ayuda de un conocimiento detallado sobre la estructura de la proteína, su función y mecanismo catalítico. Otro método relacionado es la evolución dirigida, que consiste en ciclos repetidos de mutagénesis al azar y/o recombinación genética, seguidos de un screening de alto rendimiento o selección de los mutantes funcionalmente mejorados. Este método tiene como ventaja que no se necesita información estructural de la proteína en cuestión.

1.1.10. Reacciones catalizadas por peptidasas vegetales

En la aplicación de peptidasas en procesos de síntesis a nivel industrial no existen protocolos para validar ni escalar los procesos desarrollados, y en cada caso deben optimizarse un gran número de variables (Guzmán *et al.*, 2007).

Entre las peptidasas más citadas dentro del terreno de la biocatálisis figuran subtilisina, quimotripsina, termolisina, tripsina y papaína, ya que son peptidasas muy empleadas en procesos industriales y sus propiedades bioquímicas y fisicoquímicas, así como sus estructuras, son muy conocidas. Por otro lado, están disponibles comercialmente en grandes cantidades y no son costosas.

1.1.10.1. Síntesis de péptidos

La aplicación de proteasas a la síntesis de péptidos, aunque posee varias ventajas, presenta algunas características desfavorables. Cuando la síntesis tiene lugar en solventes orgánicos, la reacción está generalmente limitada a la síntesis de di- y tripéptidos (Guzmán *et al.*, 2007). Varios ejemplos incluyen a los precursores de edulcorantes como aspartamo (Nakaoka *et al.*, 1998) y alitamo (L-Asp-D-Ala) (Kim & Shin, 2001). También fue sintetizado otro precursor de un péptido amargo, Z-Ala-Phe-OMe, empleando papaína y otra peptidasa vegetal, funastraína, depositadas en poliamida empleando solventes orgánicos con baja actividad acuosa (Morcelle *et al.*, 2006). Peptidasas vegetales han sido empleadas también en la síntesis de precursores o derivados peptídicos de interés como agentes

terapéuticos. La síntesis del derivado Bz-Tyr-Arg-OH (derivado de la kiotorfina, péptido analgésico) fue obtenida con 42% de rendimiento empleando PEG-papaína como biocatalizador (Lee *et al.*, 1993) en un sistema monofásico conteniendo 40% de etanol bajo control cinético.

La síntesis de oligopéptidos más largos mediada por catálisis enzimática es más compleja y requiere de más etapas. Estas estrategias combinan el empleo de distintas proteasas según el enlace peptídico que deba formarse, así como desprotecciones de los grupos amino terminales mediante reducción química. Así se logró la síntesis enzimática completa de derivados aminoácidos de encefalina empleando distintas combinaciones de enzimas, medios de reacción y sustratos (Clapés *et al.*, 1995; Klein & Cerovsky, 1996).

Una aplicación interesante consiste en la obtención de poliaminoácidos catalizada por peptidasas. Los poliaminoácidos son materiales poliméricos biodegradables con aplicación en medicina y cosmética. Los clorhidratos de los ésteres de Met, Phe, Thr y Tyr fueron polimerizados en buffer acuoso por acción de papaína, resultando en poli (α -aminoácidos) insolubles en agua y en la mayor parte de los solventes orgánicos. La aparición de esta clase de polímeros insolubles ha sido observada también mediante la acción de papaína y funastraína en solventes orgánicos sobre Phe-OMe (Morcelle *et al.*, 2006).

1.1.10.2. Síntesis de isómeros peptídicos

Este tipo de reacciones contemplan la capacidad de las peptidasas de acilar compuestos distintos a aminoácidos, así como emplear dadores de acilo diferentes a los aminoácidos codificados. Para que estas reacciones sean posibles, la peptidasa debe ser altamente flexible y poco específica. En general, aún en estos casos se ve que los rendimientos caen con el aumento de grado de modificación de la cadena aminoácida. Es así que solamente el acoplamiento de aceptores de acilo cercanamente relacionados con los residuos aminoácidos tiene relevancia práctica (Bordusa, 2002).

1.1.10.3. Síntesis de péptidos miméticos de RGD

En los últimos años muchos investigadores trataron de sintetizar el tripéptido RGD y péptidos que contienen la secuencia RGD por métodos químicos o enzimáticos. En la sección 1.1.8. ya hemos discutido los beneficios de la síntesis enzimática frente a los métodos químicos, y el problema que conlleva la síntesis de péptidos que contienen aminoácidos hidrofílicos, tales como el tripéptido RGD, que contiene dos aminoácidos cargados (Arg y Asp) y uno neutro (Gly) (véase 1.1.9.). Un método disponible para superar la dificultad que significa la baja solubilidad de los sustratos en disolventes orgánicos es utilizar micelas reversas como medios de reacción. Se ha encontrado que la síntesis de dipéptidos precursores del tripéptido RGD catalizada por proteasas bajo control termodinámico y cinético en micelas inversas se han llevado a cabo con un rendimiento razonable (Chen, *et al.*, 1998). Sin embargo, cabe señalar que la presencia de moléculas de surfactante en el sistema de reacción hace que la separación y purificación del producto sean difíciles. En el año 1991, Chen *et al.* informaron que una proteasa alcalina industrial, la alcalasa, obtenida de una cepa de *Bacillus licheniformis*, presentó alta estabilidad (vida media mayor a 5 días) en etanol ó 2-metil-2-propanol y fue utilizada por estos investigadores en ensayos de síntesis mediante control cinético. Por su parte Hou *et al* (2005) utilizaron la alcalasa, para lograr la unión de Bz-Arg-OEt a Gly-Asp-(NH₂)₂, mediante el control cinético de la reacción. Estos investigadores emplearon mezclas orgánico:acuosas y buffer Tris-HCl 0,1M pH 8, reportando que la proporción 85:15 fue la óptima y el máximo rendimiento fue obtenido en el medio que contenía etanol como solvente orgánico. Es notable el hecho de que este grupo observó que cuando el tiempo de reacción fue mayor al tiempo óptimo (8 horas), el rendimiento del producto se mantuvo relativamente constante, por lo que dedujeron que la hidrólisis no tendría lugar.

1.2. Introducción a las Relaciones Cuantitativas Estructura-Propiedad (QSPR) y las Relaciones Cuantitativas Estructura-Actividad (QSAR)

1.2.1. Definiciones y fundamentos

Teniendo en cuenta que las propiedades fisicoquímicas y biológicas de un compuesto químico dependen de su estructura molecular, al observar una serie de compuestos es válido preguntarnos si puede establecerse una relación sistemática entre alguna de sus propiedades y su estructura. Y suponiendo que así fuera ¿podría describirse esa relación en términos cuantitativos? En otras palabras... ¿podría formularse dicha relación como una función matemática?

Una relación estructura-actividad (SAR) se establece cuando un conjunto de propiedades de una serie de compuestos explica su actividad o respuesta biológica (ecuación 1):

$$\text{Actividad Biológica} = f(\text{características estructurales}) \quad (1)$$

Las relaciones cuantitativas estructura-actividad (Quantative Structure-Activity Relationships, QSAR) o estructura-propiedad (Quantitative Structure-Property Relationship, QSPR) reflejan en términos matemáticos esa relación. Los estudios QSPR-QSAR combinan Estadística y Química Computacional, con el objeto de establecer modelos matemáticos que cuantifiquen las relaciones SAR. Un modelo QSPR-QSAR es una hipótesis químico-matemática que tiene su fundamento en el hecho de que la estructura de una molécula es responsable de sus propiedades químicas, fisicoquímicas y biológicas (incluida su actividad farmacológica) (King, 1983; Sexton, 1950; Hansch, 1969). La teoría QSPR-QSAR puede emplearse para describir a la propiedad en términos estructurales proveyendo de una base racional para la búsqueda racional de nuevos fármacos, a través del diseño y optimización molecular, y del cribado virtual (Seydel, 1985; Topliss, 1983)

1.2.2. Técnicas QSAR en el desarrollo de nuevos fármacos

En las últimas décadas en el área de diseño de fármacos se ha impuesto la búsqueda racional de nuevos agentes quimioterapéuticos. El diseño de fármacos asistido por computadora combina la capacidad de cálculo de los ordenadores modernos con herramientas de la Estadística y los avances en el conocimiento de las bases moleculares de las distintas patologías.

En este marco, el desarrollo de modelos químico-matemáticos QSAR/QSPR se ha constituido en una herramienta de gran potencial para planificar y optimizar racionalmente los procesos de síntesis y evaluación biológica de compuestos químicos (Hansch *et al.*, 1991; Dudek *et al.*, 2006). En este sentido la metodología QSAR propone encontrar una relación entre la actividad biológica de una serie de compuestos y un conjunto de “parámetros”, denominados descriptores moleculares, que caracterizan algún aspecto de la estructura molecular de los mismos (transformando información química en variables continuas o discretas que pueden ser utilizadas como variables independientes de un modelo matemático). Se trata de inferir, a partir de una muestra de compuestos, modelos matemáticos generales que describan la relación entre la estructura molecular y una propiedad biológica determinada. Se requiere por lo tanto de un conjunto de compuestos, el conjunto de entrenamiento o calibración, para los cuales se ha determinado experimentalmente el valor de la propiedad biológica de interés. En el modelo matemático que se infiere a partir del conjunto de entrenamiento, la variable dependiente será entonces la propiedad biológica de interés y las variables independientes serán los descriptores moleculares de este conjunto de calibración. El objetivo del modelado QSAR será encontrar una función químico-matemática (ecuación 2) que vincule la variación de la actividad biológica con la estructura molecular de los compuestos que conforman el conjunto de entrenamiento:

$$\text{Propiedad} = f(\text{descriptor molecular 1, descriptor molecular 2, ..., descriptor molecular n}) \quad (2)$$

Una vez desarrollado y validado el modelo, será posible utilizarlo en la estimación de la propiedad de interés para compuestos no incluidos en el conjunto de calibración que pertenezcan al dominio de aplicabilidad (vease 1.2.5.1.) del modelo generado, de manera que el modelo podrá utilizarse para guiar racionalmente la síntesis química y/o la evaluación farmacológica de un compuesto, optimizando el uso de los recursos humanos y materiales.

1.2.3. Descriptores moleculares

Bajo el término Descriptores moleculares se agrupan una serie de parámetros que a través de valores numéricos representan una información estructural (por ejemplo electrónica o estérica) (Arranz, Velar & Garcia, 2008). Dado que las propiedades fisicoquímicas de un compuesto también dependen de su estructura molecular, las mismas también podrán utilizarse como descriptores moleculares en análisis QSAR.

Los descriptores moleculares pueden clasificarse en (Arranz, Velar & Garcia, 2008):

- Índices basados en propiedades químico-físicas: Son aquellos que derivan de la determinación experimental o de la estimación teórica de una determinada propiedad químico-física de la molécula. Podemos considerar ejemplos tales como el logaritmo del coeficiente de partición, la refractividad molecular, la constante de ionización y el momento dipolar.

- Índices mecánico-cuánticos: Son aquellos que se determinan por cálculos mecánico-cuánticos. Algunos ejemplos de este tipo de descriptores son la carga neta sobre un átomo A, las cargas netas más negativa y más positiva de la molécula y las energías HOMO y LUMO.

- Índices topológicos y topográficos: Son aquellos que se determinan a partir del grafo químico, es decir, a partir de la representación bidimensional de la estructura del compuesto. Por ejemplo, los índices de conectividad.

Por otro lado, los descriptores moleculares pueden clasificarse según dependan (descriptores 3, 4 y 5D) o no (descriptores 0, 1, y 2D) de la conformación y la orientación de la molécula en el espacio (Talevi, 2007).

Finalmente, también podemos diferenciar entre descriptores experimentales y descriptores teóricos (Talevi, 2007). Por ejemplo, distintas propiedades determinadas experimentalmente (tiempo de retención en un sistema cromatográfico determinado, punto de fusión de un compuesto químico, coeficiente de reparto entre un dado solvente orgánico y agua) pueden utilizarse como descriptores moleculares; asimismo, existen descriptores cuyo cálculo no implica una determinación experimental sino un cálculo teórico (por ejemplo en el mencionado caso de los descriptores mecano-cuánticos o cuando una propiedad fisicoquímica es estimada teóricamente mediante un modelo QSPR y luego utilizada a su vez como descriptor molecular en otro modelo).

1.2.3.1. Descriptores de baja dimensionalidad

Su valor no depende de la conformación ni de la orientación tridimensional de la molécula. Este grupo engloba varios los descriptores 0D y 1D y 2D (Talevi, 2007).

Descriptores 0D o constitucionales: Ejemplos de estos descriptores son el peso molecular, el número de átomos que componen la molécula y el número de dobles o triples enlaces incluidos en la estructura.

Descriptores 1D: Son aquellos que indican presencia, o frecuencia, de un grupo funcional, subestructura o fragmento en la molécula

Descriptores 2D o topológicos: Son aquellos que dependen de la topología molecular, es decir los que dependen de las posiciones y las interconexiones entre los átomos de la molécula.

1.2.3.2. Descriptores de alta dimensionalidad

Estos incluyen descriptores 3D. Sus valores numéricos dependen de la conformación y/o de la orientación tridimensional de la molécula.

Descriptores 3D: Los descriptores tridimensionales toman en consideración la geometría 3D de la estructura: distancias de enlaces, ángulos de enlaces, ángulos diedros, etc., pudiendo describir entonces las propiedades estereoquímicas de las moléculas. Para el cálculo de cualquier tipo de descriptor 3D se debe contar con una conformación molecular probable, para lo cual generalmente se utiliza algún método de optimización apropiado (véase 1.2.4.3.).

Adicionalmente, las aproximaciones QSAR 4D y 5D muestrean, para realizar el análisis, distintas conformaciones y estados de protonación de los ligandos y del receptor, respectivamente.

1.2.4. Clasificación de las técnicas QSAR

En la actualidad existen diferentes técnicas por las que se puede desarrollar un estudio QSAR. Según la naturaleza de la variable dependiente considerada (Talevi, 2007), los modelos QSAR pueden dividirse en:

- cualitativos o clasificatorios
- cuantitativos

La primera clase considera una variable dependiente discreta, categórica; la propiedad observada es representada mediante una etiqueta (tóxico y no tóxico; activo e inactivo; etc.). Ejemplos de métodos clasificatorios son el análisis discriminante o los árboles clasificatorios. Los modelos cuantitativos, en cambio, utilizan una variable dependiente continua y se usan para predecir, por lo tanto, el valor numérico de una propiedad determinada (por ej. $\log IC_{50}$)

Por otro lado, en relación al tipo de tratamiento matemático utilizado para cuantificar o ponderar la contribución de cada descriptor molecular a la actividad biológica podemos distinguir entre métodos lineales (regresión lineal múltiple, cuadrados mínimos parciales) o no lineales (redes neuronales) (Arranz, Velar & Garcia, 2008). No se profundizará en la descripción de las metodologías no-lineales, dada su complejidad y debido a que no serán utilizadas en el presente trabajo.

Independientemente del tipo de análisis matemático utilizado, también podemos clasificar las metodologías QSAR de acuerdo a la dimensionalidad de los descriptores moleculares utilizados.

En el presente trabajo se realizó un modelo 3D-QSAR cualitativo, mediante métodos lineales, por lo que en las siguientes secciones dedicaremos unas líneas a este tipo de aproximación en particular.

1.2.4.1. QSAR tridimensional (3D-QSAR).

Los métodos 3D QSAR más populares, entre los que se puede destacar el CoMFA (Análisis Comparativo de Campos Moleculares) (Cramer *et al.*, 1988), se basan en disponer cada molécula en una grilla y calcular el valor de la fuerza que diferentes campos (estéricos y electrostáticos) ejercen sobre los vértices de la misma, estableciendo qué campo, y en qué zona de la grilla, es significativo para la actividad biológica (Fig. 1.8). La premisa básica de las metodologías basadas en grilla, es que el muestreo adecuado de los campos alrededor de un conjunto de moléculas o fármacos, puede proporcionar la información necesaria para comprender la actividad biológica.

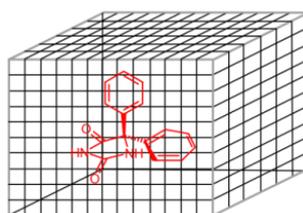


Figura 1.8: Las metodologías 3D-QSAR basadas en grillas. En esta figura se muestra una representación de una grilla en cuyo interior se encuentra una molécula de difenilhidantoína.

De lo mencionado anteriormente, se desprende que en las metodologías tipo CoMFA deberá definirse no sólo una conformación probable de las estructuras utilizadas para el análisis (requisito indispensable para el correcto cálculo de los descriptores 3D) sino también un criterio de orientación de las mismas en la grilla. Las metodologías 3D que no se basan en grillas en cambio, sólo exigen la primera de esas dos condiciones, esto es, la

búsqueda de un conformero de baja energía (Talevi, 2007). Esta será la aproximación que utilizaremos en el presente trabajo.

1.2.5. Etapas generales de un estudio QSAR

En general, un estudio QSAR pueden dividirse en las siguientes etapas (Talevi, 2007):

1. Selección del conjunto de entrenamiento.
2. Cálculo del valor, para los compuestos del conjunto de entrenamiento, de los D descriptores que se utilizarán en el estudio.
3. Selección de un subconjunto de d descriptores (a partir del pool de D posibles variables independientes) que mejor caracterizan la propiedad estudiada.
4. Validación.
5. Aplicación para la predicción de casos no incluidos en el conjunto de entrenamiento.

1.2.5.1. Generación del conjunto de entrenamiento.

Se denomina conjunto de entrenamiento al conjunto de estructuras químicas a partir del cual se deriva el modelo QSAR. Este conjunto define el “espacio químico” dentro del cual la aplicación del modelo tendrá validez. En otras palabras, el valor de la actividad/propiedad biológica predicho por el modelo para un compuesto externo al conjunto de entrenamiento será confiable si y sólo si el mismo posee cierta similitud con alguna o varias de las estructuras incluidas en el conjunto de entrenamiento. El espacio químico dentro del cual la aplicación del modelo QSAR es confiable se denomina “dominio de aplicabilidad”.

La calidad de los modelos obtenidos dependerá fuertemente de la calidad de la serie de compuestos empleados como conjunto de entrenamiento. En consecuencia, es deseable que el conjunto de estructuras utilizadas como conjunto de entrenamiento cumpla los siguientes requisitos:

Diversidad. Los productos de la serie deben ser, respecto a los parámetros elegidos, diferentes entre sí. Además, puesto que la serie debe ser una muestra del espacio experimental, el rango de variación de las características dentro de dicha serie debe ser equiparable a la que existe en el espacio experimental.

Mismo mecanismo de acción. Esto es, los compuestos del conjunto de entrenamiento deben interactuar con el mismo receptor o blanco molecular. En nuestro caso, todos los compuestos deben ser RGD miméticos que interactúen con receptores del tipo de las integrinas.

Homogeneidad en la determinación del dato experimental. En el desarrollo de modelos cuantitativos sería deseable que los valores de la variable dependiente hayan sido determinados experimentalmente a través del mismo ensayo e idealmente en el mismo laboratorio, por el mismo laboratorista. De esta forma el modelo sólo reflejaría la relación entre la variabilidad estructural y la variabilidad en la respuesta, y no otras variabilidades que estén relacionadas con diferencias en la determinación experimental de los diferentes compuestos del conjunto de entrenamiento.

Los modelos clasificatorios, en este sentido son más flexibles, ya que se asignan valores discretos de la variable dependiente a los compuestos del conjunto de entrenamiento, por lo que gran parte de la variabilidad introducida por variaciones experimentales (que genera ruido en el modelo) es absorbida por el esquema de clasificación. Para la construcción de este tipo de modelos, no importa el valor exacto de la actividad farmacológica; sólo interesa que la actividad de un compuesto determinado esté por encima o por debajo de un valor de corte, es decir, si el compuesto es o no activo según un criterio definido por el modelador. Por ello, sólo aquellos compuestos que posean una respuesta cercana al punto de corte considerado pueden introducir cierto error en el modelo. De lo expuesto anteriormente se desprende que cuando el conjunto de entrenamiento es generado a partir de la recopilación de estructuras mediante una búsqueda bibliográfica, como se realizó en este trabajo, el desarrollo de modelos QSAR clasificatorios es más fiable y correcto que obtener un modelo cuantitativo que se crea que pueda tener error debido a la

alta variabilidad en el resultado de la determinación experimental de la respuesta biológica (Penzotti *et al.*, 2002).

Buena distribución de valores de la variable dependiente. Los valores observados de la propiedad estudiada, deberían presentar una buena distribución para los compuestos del conjunto de entrenamiento. La importancia de este punto se puede comprender si se considera el siguiente ejemplo. Supongamos que el conjunto de entrenamiento esta esencialmente formado, por ejemplo, por compuestos de alta y mediana actividad. En este caso el modelo obtenido sólo podrá predecir con exactitud valores de compuestos muy activos y moderadamente activos, pero no identificará qué características estructurales son francamente desfavorables a la actividad por no estar los compuestos poco activos e inactivos representados dentro del conjunto de entrenamiento.

1.2.5.2. Cálculo del valor de un conjunto de D descriptores para el conjunto de entrenamiento.

Lo fundamental en esta etapa es chequear las estructuras del conjunto de entrenamiento y, en los casos en que sea necesario (3D-QSAR y metodologías QSAR de mayor dimensionalidad) realizar un pre-procesamiento adecuado de las estructuras de los compuestos, lo que permita calcular los descriptores moleculares dependientes de la conformación a partir de una conformación razonable. Usualmente se define la conformación más probable como la conformación de mínima energía, es decir, el mínimo global de energía en el espacio conformacional. Este proceso de optimización de la estructura requiere la implementación de métodos teóricos de cálculo a fin de obtener la geometría molecular de mínima energía. Aún para moléculas relativamente pequeñas (con ocho o más enlaces con libre rotación) la exploración exhaustiva del espacio conformacional en búsqueda del mínimo absoluto de la hypersuperficie de energía no es viable (demandaría un costo/tiempo computacional prohibitivo). En general, la optimización de geometría se hace entonces con un algoritmo de optimización local, lo que implica que puede converger a mínimos locales diferentes dependiendo de la geometría inicial de partida. Estos métodos de minimización son generalmente métodos descendentes, es decir que va descendiendo por

la superficie o hipersuperficie de energía potencial hasta alcanzar el mínimo local más próximo. Son procedimientos iterativos mediante los cuales a partir de un valor inicial x_0 (coordenadas de la conformación inicial o de partida) se va pasando por diferentes valores hasta alcanzar el x_m (coordenadas del mínimo local al que se llega aplicando un método de optimización determinado a determinada conformación inicial). El mínimo al que se llega dependerá de la conformación inicial, la dirección de búsqueda y el algoritmo de minimización empleado. La dependencia de las conformaciones iniciales implicará la necesidad de explorar la hipersuperficie de energía a partir de distintas conformaciones iniciales, para lo cual se podrá recurrir a una variación sistemática de los ángulos dihedros de las conformaciones iniciales, o a un método estocástico (descrito en el capítulo número 3). La dirección descendente en la función de energía se establece a partir del gradiente de la función utilizada para el cálculo de la energía potencial asociada a una conformación; el gradiente indicará, en cada punto de la hipersuperficie de energía, la dirección hacia la cual la función de energía crece más rápidamente. Análogamente, el negativo del gradiente indica la dirección en la cual la función de energía potencial decrece más rápidamente, esto es, cómo avanzar lo más rápido posible hacia el mínimo local.

Luego de realizar el muestreo de conformaciones iniciales y su posterior optimización para arribar a un mínimo de energía, se elige de entre todos ellos el de menor energía esperando (sin certidumbre) que corresponda al mínimo absoluto. No obstante, se debe tener presente, que no es seguro que la conformación de menor energía sea la más activa en un sistema biológico. Debe tenerse en cuenta, además, que a menudo, el fármaco experimenta cambios conformacionales inducidos por el mismo receptor.

Teniendo en cuenta la importancia de contar con la estructura optimizada, es fundamental la elección del método empleado para calcular la energía asociada a un conformero determinado. En su libro, Arranz, Velar & Garcia (2008) distinguen y discuten la implementación de dos tipos de métodos:

- ✓ los métodos basados en la Mecánica Clásica o Newtoniana, entre ellos la mecánica molecular (MM);
- ✓ y los basados en la Mecánica Cuántica, como los métodos semi-empíricos y los métodos *ab-initio*.

La elección de uno u otro depende fundamentalmente del tamaño de la molécula, la naturaleza del problema planteado y del tiempo de cálculo requerido.

La MM, es un método empírico. Es el menos exacto de todos, pero también requiere de menos tiempo de cálculo con respecto a los métodos mecano-cuánticos. La MM trata la molécula como si fuera un conjunto de átomos o puntos en el espacio gobernados por una serie de potenciales regidos por las leyes de la mecánica clásica. Se considera que los núcleos están sometidos a un campo de fuerzas que se compone de una serie de términos que modelan las tensiones de los enlaces, de sus ángulos, sus torsiones, las interacciones de Van der Walls, electrostáticas y los puentes de hidrogeno. La hipótesis de partida es que los enlaces tienen longitudes y ángulos naturales o de equilibrio a los que las moléculas se ajustan. Una desviación de estos valores implica un aumento de la energía del sistema. La energía potencial total de la molécula U (ecuación 3) se calcula como suma de energías de diferentes movimientos que están presentes en la molécula:

$$U = \Sigma U_r + \Sigma U_\theta + \Sigma U_\omega + \Sigma U_{VDW} + \Sigma U_D + \Sigma U_H \quad (3)$$

donde U_r define las deformaciones de la longitud de enlace (acortamiento-alargamiento), U_θ se relaciona con las distorsiones del ángulo de enlace, U_ω es la energía de torsión del ángulo diedro por rotación de los enlaces y U_{VDW} , U_D y U_H consideran las contribuciones de las interacciones de Van der Walls, las repulsiones electrostáticas y las asociaciones por puentes de hidrógeno, respectivamente (Arranz, Velar & Garcia, 2008). El punto más delicado de este tipo de métodos de cálculo es la parametrización de cada uno de los campos de fuerza. La parametrización consiste en tomar un conjunto de moléculas (de geometría y energía conocidas experimentalmente) e ir modificando las constantes arbitrarias

(parámetros) que aparecen en los diferentes potenciales hasta reproducir lo mejor posible la geometría, energía y otros valores de ese conjunto de moléculas tomadas como modelo. Posteriormente, una vez parametrizado el método, se podrán predecir propiedades de otras moléculas. Los campos de Fuerza Experimentales se caracterizan porque la mayoría de datos utilizados para determinar las diferentes constantes de fuerza de la expresión de la energía provienen de medidas experimentales.

Las geometrías optimizadas por métodos MM se pueden utilizar luego como datos de entrada para los cálculos mecánico-cuánticos (diremos que refinamos los conformeros obtenidos).

Por otra parte, en los métodos de cálculo basados en las leyes de la mecánica cuántica, los estados cuánticos, energía y otras propiedades relacionadas, se obtienen resolviendo la ecuación de Schrödinger. Estos métodos de cálculo, a diferencia de los métodos basados en la mecánica clásica, consideran de forma explícita los electrones. Sin embargo para muy pocos sistemas es posible resolver dicha ecuación en forma exacta. Es por esto que existen diferentes aproximaciones matemáticas que permiten resolverla (Arranz, Velar & Garcia, 2008). En párrafos anteriores hicimos mención de que existen dos métodos principales de cálculos mecánico-cuánticos:

1. Los métodos semi-empíricos
2. Los métodos *ab-initio*

Los métodos semi-empíricos, utilizan parámetros experimentales para simplificar el cálculo computacional. Estos métodos resuelven en forma aproximada la ecuación de Schrödinger que depende de parámetros apropiados al tipo de sistema químico bajo estudio. Tienen un costo computacional intermedio, y proveen una descripción cualitativa razonable.

Los métodos *ab-initio*, a diferencia de MM y métodos semi-empíricos, no utilizan parámetros experimentales en el cálculo, y están basados únicamente en las leyes de la mecánica cuántica y valores de las constantes físicas fundamentales (velocidad de la luz, masas y cargas de electrones y protones, constante de Plank, etc). La ecuación de Schrödinger del sistema se resuelve tan correctamente como se pueda, realizando el menor

número de aproximaciones. Estos métodos predicen con mayor exactitud desde el punto de vista cuantitativo y cualitativo, pero el costo computacional aumenta sustancialmente.

Una vez optimizada la molécula a través de alguno de los métodos anteriores (o de varios de ellos utilizados en serie) pueden calcularse los descriptores 3D. Para esto existen varios software disponibles, entre ellos el Dragon (DRAGON, Versión 4.0, 2003), que fue utilizado en este trabajo.

1.2.5.3. Selección y ponderación de un subconjunto de d descriptores: generación del modelo

En esta etapa, se efectúa una selección de un subconjunto de d descriptores, a partir de todo el conjunto de D posibles variables independientes calculadas en el punto anterior, a fin de encontrar los que mejor caractericen a la propiedad estudiada. En este trabajo se seleccionaran aquellos descriptores que mejor describan si una estructura es activa o inactiva.

1.2.5.3.1. Regresión Lineal

Los métodos más ampliamente utilizados para relacionar una variable dependiente de interés con un conjunto de descriptores moleculares son los métodos lineales multivariable. Entre estos métodos se destaca la Regresión Lineal Múltiple (MLR), debido a su simplicidad de cálculo y de interpretación de los resultados (Talevi, 2007; Arranz, Velar & Garcia, 2008).

La regresión lineal múltiple se basa en el método de los cuadrados mínimos, esto es, encontrar la combinación lineal de variables que minimice la suma de los cuadrados de los residuales, definidos como la diferencia entre los valores observados de la variable dependiente y los valores predichos. El resultado que se obtendrá es la ecuación de una recta, un plano, o de un hiperplano, según sea el número de variables independientes incluidos en la expresión. En este método, una vez establecidos los conjuntos de valores de las variables independientes X_i y la actividad biológica A , se obtiene un modelo en forma de ecuación de una recta (ecuación 4), la cual describe la dependencia de la actividad (A) en

función del conjunto de descriptores (X_i) así como la magnitud de las contribuciones de cada uno de ellos.

$$A = f(X_i) + \text{constante} \quad (4)$$

El modelo así obtenido ha de ser analizado en función de su calidad estadística, para poder evaluar su capacidad de predicción, debido a que cuanto mayor sea su calidad, más confiables y exactas serán las predicciones que se podrán realizar. Usualmente se utilizan el coeficiente de determinación (r^2) y la desviación estándar en la estimación (**SEE**), para evaluar la bondad del ajuste de los valores de la variable independiente predichos por el modelo con respecto a los valores observados. El coeficiente de determinación indica la fracción de la varianza total de los datos experimentales que puede ser explicado por el conjunto de descriptores incluidos en el modelo, mientras que la desviación estándar depende de la varianza no explicada por las variables independientes y de los grados de libertad del modelo, y es una medida de cuánto se alejan los valores predichos por el modelo de la línea, plano o hiperplano. La tendencia a cero de este valor pudiera presuponer mayor calidad en la predicción. Sin embargo, esto puede conducir a modelos sobrepredictivos en los cuales se refleje exactamente el comportamiento de la muestra (es decir, del conjunto de entrenamiento) pero que no sea posible utilizarlo para la extrapolación de valores en el caso de la predicción de nuevos compuestos. A este fenómeno se lo conoce como sobre-ajuste (véase 1.2.5.3.3.) o “paradoja de Kubinyi”: muy alta capacidad de predicción interna (conjunto de entrenamiento) suele resultar en baja capacidad de predicción externa (conjunto de prueba y población general) (Van Drie, 2003; Golbraikh & Tropsha, 2002). Un valor de SEE es válido cuando está en el mismo orden de magnitud del error experimental de las mediciones de la variable dependiente (es decir, la propiedad o actividad biológica).

Por otro lado, en todo análisis de MLR debe verificarse que:

- Los residuales estén normalmente distribuidos alrededor de una media de 0.
- El número de variables independientes sea menor que el número de casos.
- No existe correlación lineal significativa entre las variables independientes. En general, se limita a cierto rango de valores al coeficiente de correlación para cualquiera de los pares de descriptores incluidos en el modelo. Otra forma de restringir el ingreso de pares de descriptores altamente correlacionados en el modelo es fijar una cota a la tolerancia β definida como $1 - r^2$, donde r^2 es en este caso el coeficiente de determinación entre una de las variables independientes del modelo y todas las demás variables independientes.

Las técnicas de selección de variables independientes más utilizadas en conjunto con la MLR son las conocidas como Forward Stepwise y Backward Stepwise.

En este trabajo se utilizó la metodología Forward Stepwise, que consiste en llevar a cabo una regresión por cuadrados mínimos por pasos. En cada etapa, o paso, la variable independiente del pool de D variables mejor correlacionada con la variable dependiente, es incluida en la ecuación de regresión; en el segundo paso, se adiciona la variable independiente que de entre todas las remanentes produce un incremento mayor en el r^2 . Este proceso continúa hasta que la adición de una nueva variable no contribuye a un aumento estadísticamente significativo del r^2 , o hasta que se haya realizado el número de pasos solicitados por el modelador. Un test de Fisher parcial puede ser utilizado para corroborar si la inclusión de una variable en un modelo induce un aumento significativo de r^2 . El valor F correlaciona la varianza explicada (r^2) por el número de grados de libertad con la varianza no explicada ($1 - r^2$) por el número de variables del modelo. Cuanto más alto es el porcentaje de varianza explicada por el modelo, mayor será el valor de F , mientras que la existencia de variables independientes con baja contribución a la explicación de la varianza, tenderán a disminuir dicho valor. Por otro lado, al aumentar el número de casos (n), el valor numérico de este estadístico se incrementa; lo inverso ocurre al incrementarse el número de descriptores incluidos en el modelo (d), ya que F es inversamente proporcional a d , y tiende a cero cuando d se aproxima a n . Esto puede comprenderse si se considera que a medida que aumenta el número de variables incluidas en el modelo, aumenta la probabilidad de

correlación azarosa, esto es, aumenta la chance de que alguna de las variables independientes remanentes (descriptores no incluidos) refleje la varianza aun no explicada por los descriptores que sí están incluidos en el modelo.

Alternativamente, la metodología Stepwise Backward comienza incluyendo todas las variables del pool en el modelo y elimina, una a una, variables independientes hasta que la eliminación de una variable reduce significativamente el r^2 .

1.2.5.3.2. Análisis Lineal Discriminante

El Análisis Lineal Discriminante (ALD) para dos grupos o Análisis Lineal Discriminante de Fisher puede pensarse (y es análogo a) una regresión lineal múltiple. Si se codifican los dos grupos a través de una variable binaria (de valores 1 y -1, por ejemplo) y se utiliza esta variable binaria en un análisis MLR, obtendremos una ecuación lineal o función discriminante (**df**) (ecuación 5), capaz de distinguir entre los elementos de un grupo u otro (Electronic Statistics Textbook, 2011).

$$\mathbf{df} = \mathbf{a}_0 + \mathbf{a}_1\mathbf{d}_1 + \mathbf{a}_2\mathbf{d}_2 + \dots + \mathbf{a}_n\mathbf{d}_n \quad (5)$$

donde en nuestro caso, **df** indica si un compuesto es activo o inactivo, $\{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n\}$ se refiere –una vez más– a un conjunto de descriptores capaces de discriminar las clases o categorías propuestas y $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ es el conjunto de coeficientes de regresión.

El estadístico que caracteriza este tipo de análisis es el λ de Wilks, definido como la relación entre el determinante de la matriz varianza/covarianza dentro de los grupos sobre el determinante de la matriz varianza/covarianza total. Este estadístico es una medida directa de la proporción de la varianza explicada por la variable clasificadora (es decir, evalúa la relación entre la varianza entre grupos y la varianza total). Cuando este parámetro toma el valor cero, la discriminación de las categorías consideradas es perfecta y cuando λ es igual a uno, la capacidad de separación de las clases es nula. Esencialmente, el análisis discriminante puede interpretarse como una generalización del análisis de varianza (ANOVA) (utilizado para definir si dos grupos son diferentes entre sí respecto a la media

que una variable determinada asume para cada grupo) cuando lo que se desea es discriminar entre grupos a partir de un conjunto de variables.

Al igual que en la regresión lineal múltiple, en este tipo de análisis las técnicas de selección de variables independientes más utilizadas son Forward Stepwise y Backward Stepwise.

1.2.5.3.3. Sobre-ajuste

En estadística, el sobre-ajuste o sobre-entrenamiento se produce cuando un modelo estadístico describe el error aleatorio o ruido en vez de la relación subyacente entre la variable dependiente y las independientes. El sobre-ajuste generalmente ocurre cuando el modelo es demasiado complejo o demasiado flexible. Cuando el modelo contiene muchos parámetros en relación con el número de observaciones, éste comienza a memorizar los datos de entrenamiento en lugar de aprender a generalizar a partir de ellos. Como ejemplo extremo considérese un modelo cuyo número de parámetros sea igual o mayor que el número de observaciones. Este modelo podrá predecir perfectamente los datos de entrenamiento con sólo memorizar todos los datos de entrenamiento pero el error que se observará al utilizar el modelo para predecir el valor de la variable dependiente para otro compuesto será inaceptable. Los modelos “sobre-ajustados” generalmente tienen rendimientos predictivos pobres, ya que pueden exagerar las fluctuaciones de menor importancia en los datos (Babyak, 2004).

Existe una regla empírica (Babyak, 2004; Peduzzi *et al.*, 1996; Peduzzi *et al.*, 1995), que establece que el riesgo de sobre-ajuste disminuye si se cuenta con entre 10 y 15 casos por variable independiente utilizada en el modelo.

Por otro lado, si bien se ha demostrado ampliamente la utilidad de esta regla, siempre es necesaria la validación del/los modelo/s obtenido/s.

1.2.5.4. Validación.

El proceso de validación es una etapa crítica y tiene por objeto la determinación de la robustez y capacidad predictiva del modelo.

Podemos distinguir dos clases generales de validación:

- ~ la validación interna
- ~ la validación externa

1.2.5.4.1. Validación Interna

Durante el proceso de validación interna, se reutiliza el mismo conjunto de entrenamiento para validar el modelo generado. Se pueden utilizar dos aproximaciones diferentes: la validación cruzada y ensayo de aleatorización de Fisher (Yasri, *et al.*, 2001).

La aleatorización de Fisher está destinada a minimizar la probabilidad de correlación fortuita entre los descriptores y la variable dependiente. El mismo se lleva a cabo “mezclando” al azar los valores observados de la variable dependiente entre los n compuestos, obteniendo un nuevo modelo con la variable dependiente aleatorizada. En otras palabras, se destruye la correlación entre los descriptores y la variable dependiente. Este procedimiento se suele repetir tantas veces como compuestos tenga el conjunto de calibración. Idealmente, el modelo original será estadísticamente muy superior a los modelos aleatorizados.

La validación cruzada leave-group-out consiste en excluir un conjunto de x compuestos del conjunto de entrenamiento y recalculando el modelo, utilizando aquellos descriptores que constituyen el modelo original (es decir, aquel generado con el total de las estructuras del conjunto de entrenamiento). El nuevo modelo así obtenido se utiliza para predecir el valor de la variable dependiente de los compuestos que habían sido excluidos.

1.2.5.4.2. Validación Externa

La validación externa consiste en predecir el valor de la variable dependiente para un conjunto de compuestos que no han sido utilizados para generar el modelo y cuya actividad se conoce. El principal problema de este proceso es que deben resignarse casos que podrían utilizarse en la generación del modelo para evaluar su predictividad, lo cual no es viable cuando se cuenta con pocas estructuras con valor experimental conocido de la propiedad

biológica de interés. Algunos autores sugieren que entre un 10 y 20% del total de estructuras disponible es suficiente (Yasri, A.; 2001), y la gran mayoría de los especialistas reconoce a la validación externa como etapa imprescindible (Golbraikh A & Tropsha A, 2002), aunque existen diferentes posturas acerca de la importancia relativa de ambas validaciones, e incluso sobre la cantidad de estructuras que debieran utilizarse al momento de realizar la validación externa.

1.2.5.5. Interpretación y aplicación del modelo a la predicción de casos no incluidos en el conjunto de entrenamiento.

Luego de haber generado los modelos y validarlos, aún queda interpretarlos. Como en todo modelo lineal, hay dos componentes fundamentales: las variables independientes, y los coeficientes de regresión.

En los modelos QSAR, las variables independientes son los descriptores, y el análisis del significado que tienen aquellos descriptores que constituyen el modelo nos da una idea del significado físico-químico del modelo, ya que estos reflejan determinadas “características moleculares”. Además la magnitud en la cual cada una de estas características contribuye, es reflejada en el valor del coeficiente (normalizado) que las acompaña.

Por otro lado, al aplicar el modelo a compuestos que no formen parte del conjunto de entrenamiento es fundamental tener en cuenta el dominio de aplicabilidad dentro del cual es confiable la predicción. Desde luego, la verificación experimental (mediante ensayos biológicos en el caso de predecir bioactividades) constituye la prueba definitiva de la validez del modelo.

2. Objetivos

2.1. Objetivos generales

- ~ Adquirir entrenamiento en la búsqueda de material bibliográfico.
- ~ Obtener adiestramiento en el manejo de procedimientos básicos de laboratorio: preparación de soluciones buffer y otros reactivos de uso común; cuantificación de proteínas; determinación de la actividad enzimática empleando distintos sustratos; preparación de biocatalizadores; desarrollo de distintas técnicas cromatográficas.
- ~ Desarrollar criterios para analizar y discutir resultados experimentales.
- ~ Integrarse adecuadamente a un grupo de trabajo.

2.2. Objetivos específicos

- ~ Obtención y caracterización del extracto con actividad proteolítica proveniente del látex de frutos de la enredadera nativa *Araujia hortorum* Fourn. (Asclepiadaceae).
- ~ Síntesis de derivados peptídicos y peptidil amino alcoholes a partir de la condensación entre el éster etílico de N^α -benzoil-L-arginina y diversos nucleófilos empleando cisteín peptidasas de *A. hortorum* como catalizador, en distintas configuraciones (enzima libre e inmovilizada en soportes inertes).
- ~ Desarrollo y validación de un modelo QSAR-3D capaz de identificar análogos sintéticos de la secuencia peptídica RGD, reconocida por receptores de integrinas.

- ~ Aplicación del modelo QSAR-3D desarrollado para predecir si los compuestos obtenidos mediante síntesis enzimática son activos o inactivos sobre receptores del tipo integrinas.

3. Metodología

3.1. Síntesis enzimática de derivados de N^α -benzoil arginina (Bz-Arg)

3.1.1. Material vegetal

Como material de partida para obtener las enzimas, se empleó el látex proveniente de frutos de *Araujia hortorum* Fourn.

Araujia hortorum, de la familia Asclepiadaceae, es una enredadera trepadora perenne con tallos amarecados esparcidos y desordenados nativa de América del Sur, conocida vulgarmente como “tasi” o “doca”. Sus hojas simples y simétricas, de forma ovado triangular, pecioladas, obtusas en la base y discolores (oscuras y glabras -carente de pilosidades- en el haz; albotomentosas -blanquecina y cubierta de pilosidades cortas- en el envés), tienen una longitud que varía entre los 4 y 9 centímetros. Sus flores se caracterizan por ser perfumadas, en forma de corola y de color blanco o rosado, con apéndice membranáceo en la parte superior y el tubo de la corona soldado a la corola; tienen una longitud promedio de 15 milímetros, siendo la época de floración desde finales de primavera hasta inicios del otoño. Los frutos, folículos ovoideos, lisos, glabros, con una longitud promedio de 14 centímetros, son de color verde azulado, tornándose marrones y de textura amaderada a medida que maduran, abriéndose finalmente para liberar las semillas. Estas últimas son de color negro, miden aproximadamente 4 milímetros de largo y poseen en uno de sus extremos pelos blancos de textura sedosa y 2,5 centímetros de longitud, los cuales facilitan la dispersión de las mismas por acción del viento (Obregón, 2008; Australian Weeds Committee, 2011). Posee un látex lechoso que brota de los frutos cuando éstos son dañados o cortados.

3.1.2. Obtención de la preparación enzimática

Se colectaron veinte frutos de *A. hortorum* de plantas crecidas en la localidad de Tolosa, partido de La Plata, en el mes de abril de 2010. Luego de lavarlos con abundante agua corriente, el látex de los mismos se recogió en una solución 0,5 mM EDTA y Na₂SO₃ 5 mM como agentes protectores. La finalidad de utilizar estos agentes protectores es la de evitar la acción de fenoloxidasas e impedir la oxidación de los grupos –SH de las proteasas cisteínicas contenidas en el látex (Priolo *et al*, 2000; Obregón *et al*, 2001).

La suspensión fue centrifugada 30 minutos a 9600 × g a 4°C a fin de eliminar las gomas y otros materiales insolubles. El sobrenadante fue centrifugado nuevamente en las mismas condiciones y el extracto crudo así obtenido (34 ml) fue fraccionado y liofilizado para su posterior utilización. El extracto crudo liofilizado y reconstituido en agua constituyó la muestra vegetal empleada para realizar los ensayos posteriores.

3.1.3. Caracterización de araujaína

El extracto proteolítico obtenido (araujaína) fue caracterizado mediante la determinación de la cantidad de proteínas, actividad caseinolítica, y estabilidad en medios orgánico:acuoso.

3.1.3.1. Determinación de la cantidad de proteína por el método de Bradford

La cantidad de proteínas presentes en el extracto fue determinada mediante el método de Bradford (Bradford, 1976). Se realizó una curva de calibración utilizando albúmina sérica bovina (BSA) (Sigma) en concentraciones de entre 100 y 1000 µg/ml. A 50 µl de cada dilución de albúmina se le agregó 2,5 ml de reactivo de Bradford y el blanco se realizó empleando 50 µl de agua destilada (AD). Se midió la absorbancia a 595 nm en un espectrofotómetro Agilent 8453 E. Cada punto de la curva se realizó por triplicado.

El mismo protocolo se aplicó a la muestra vegetal.

3.1.3.2. Determinación de actividad enzimática

La actividad proteolítica de araujaína se determinó utilizando caseína (Sigma) como sustrato (Priolo *et al.*, 1998). A 0,1 ml de la muestra se le agregó 1,1 ml de solución 1% (p/v) de caseína en buffer Tris-HCl 0,1 M de pH 8,0 conteniendo 20 mM en cisteína como activador. La reacción se llevó a cabo incubando a 37°C y se detuvo luego de dos minutos, con el agregado de 1,8 ml de ácido tricloroacético (TCA) 5% (p/v). Luego, los tubos fueron colocados en la heladera durante 30 minutos y posteriormente centrifugados durante 20 minutos a 5000 rpm. Se midió la absorbancia de los sobrenadantes a 280 nm. El ensayo blanco fue realizado inactivando la enzima con TCA 5% previo al agregado del sustrato. Cada determinación se efectuó por triplicado y se realizó el blanco correspondiente por duplicado.

La actividad proteolítica fue expresada en unidades enzimáticas arbitrarias (Unidad caseinolítica, Ucas) y en Ucas/mg de proteína, siendo 1 Uca la cantidad de enzima que produce un incremento de 1 unidad de absorbancia a 280 nm, en 1 ml de muestra, en 1 minuto de reacción, en las condiciones de ensayo (Priolo *et al.*, 2000).

3.1.3.3. Ensayos de estabilidad y actividad caseinolítica residual de araujaína

A fin de seleccionar las mezclas orgánico:acuosas más apropiadas para llevar a cabo las reacciones de síntesis, se realizaron ensayos de estabilidad y actividad caseinolítica residual de araujaína luego de haber sido incubada durante distintos tiempos, en diferentes medios de mediano/bajo contenido acuoso (Hou *et al.*, 2005).

A preparaciones de 0,1% (p/v) del extracto crudo liofilizado en buffer ácido bórico-borato de sodio 0,1 M pH 8,0 se le agregaron una cantidad suficiente de solvente orgánico tal que en todos los casos la proporción final de la mezcla orgánico:acuosa fue 85:15. Los solventes orgánicos ensayados fueron acetonitrilo (ACN), metanol (MeOH) y etanol (EtOH), y se tomaron muestras luego de 5 minutos, 1, 2, 3, 6 y 24 horas de incubación a temperatura ambiente (30°C). El ensayo control se realizó utilizando 100% de buffer.

Con el fin de optimizar la determinación de la actividad caseinolítica residual, se utilizó como muestra una preparación 0,1% (p/v) del extracto crudo liofilizado reconstituido en el mismo buffer (ácido bórico-borato de sodio 0,1 M pH 8,0). El ensayo de actividad se

realizó según lo detallado en la sección 3.1.3.2. y se ensayaron tres tiempos de reacción (2, 5 y 10 minutos). En base a los resultados obtenidos (veáse 4.1.1.3.) se eligió como tiempo 10 minutos para ensayar actividad caseinolítica residual y se consideró 100% de actividad la presentada por el control, a tiempo cero.

3.1.4. Inmovilización por adsorción

La inmovilización se realizó por deposición de araujaína en poliamida-6 (EP-700, tamaño de partícula <800 μm , tamaño medio de poro 50–300 nm, área superficial 8,4 $\text{m}^2 \text{g}^{-1}$ según el método BET, Azko) (Clapés P. *et al.*, 1999).

El polvo liofilizado resultante (100 mg) fue disuelto en 1 ml de buffer ácido bórico-borato de sodio 0,1 M pH 8 conteniendo EDTA (1 mM) y 50 mg de 1,4-ditiotreitol (DTT). La solución enzimática así obtenida fue mezclada con el soporte sólido (1 g) y la mezcla resultante fue secada al vacío (40 bar) durante 24 horas (Morcelle *et al.*, 2006).

3.1.5. Determinación de la actividad amidolítica de araujaína inmovilizada

Las unidades enzimáticas del extracto proteolítico inmovilizado fueron determinadas usando el clorhidrato de *N*^α-benzoyl-L-arginina etil éster (BAPNA, Sigma) como sustrato (Ota S. *et al.*, 1964). La enzima adsorbida (30 mg) fue suspendida en buffer fosfatos 0,1 M pH 7,4 conteniendo 1 mM de EDTA. A esta mezcla se le agregó un volumen acorde de solución stock de BAPNA (40 mM) en dimetil sulfóxido (DMSO) tal que la concentración final fue de 4 mM. La reacción se efectuó a 37°C durante 5 minutos, en agitador orbital (150 rpm) y se detuvo con el agregado de 0,5 ml de ácido acético 30% (v/v). Luego de centrifugar durante 20 minutos a 2500 \times g, se midió la absorbancia de los sobrenadantes a 410 nm (Morcelle *et al.* 2008).

Utilizando una curva de calibración de *p*-nitroanilina (producto de la reacción de hidrólisis del BAPNA) se calculó la actividad, que fue expresada en Unidades Internacionales (UI) y en Unidades Internacionales/mg de inmovilizado, siendo una UI la cantidad de enzima que produce 1 μmol de producto en 1 minuto de reacción en las condiciones del ensayo.

3.1.6. Síntesis enzimática de derivados de Bz-Arg utilizando glicinol como nucleófilo. Seguimiento por TLC y RP-HPLC.

Se implementaron dos protocolos alternativos para llevar a cabo la síntesis de Bz-Arg-Glyol (Figura 3.1).

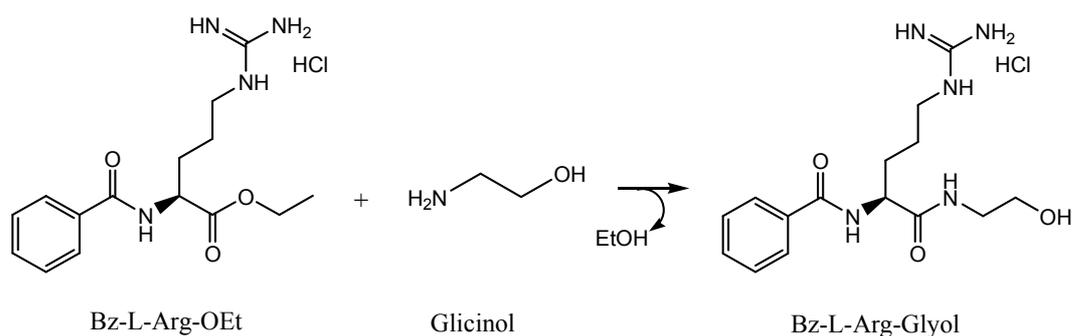


Figura 3.1: Síntesis de Bz-Arg-Glyol.

3.1.6.1. Primer protocolo de reacción: enzima inmovilizada y medio de bajo contenido acuoso

En el primer protocolo ensayado (Morcelle *et al.*, 2008), la reacción se llevó a cabo utilizando como acil donador el éster etílico de N^ε-benzoil arginina (Bz-Arg-OEt, Sigma) y glicinol (Fluka) como nucleófilo en ACN con 0,25% (v/v) de agua y 200 mg de araujiaina inmovilizada como catalizador. El volumen de mezcla de reacción utilizado fue 2 ml y las concentraciones finales de dichos reactivos en la mezcla fueron 20 mM y 30 mM respectivamente.

La reacción se realizó a 36°C en agitador orbital (120 rpm) y se tomaron muestras de 100 µl a diferentes tiempos (1, 3 y 6 hs). Se realizaron los correspondientes blancos de enzima y de sustrato, según se detalla en la tabla 3.1. La reacción se detuvo con el agregado de 20 µl de ácido acético glacial y 380 µl de MeOH.

Tabla 3.1: Concentraciones finales de los reactivos utilizados para la síntesis enzimática realizada siguiendo el primer protocolo de reacción.

Medio	Reactivo	Mezcla de reacción	Blanco de enzima	Blanco de sustratos
ACN (H ₂ O 0,25%)	Bz-Arg-OEt	20 mM	20 mM	-
	Glicinol	30 mM	30 mM	-
	araujaína inmovilizada	200 mg	-	200 mg

3.1.6.2. Segundo protocolo de reacción: enzima libre y medio de mediano contenido acuoso

En el segundo protocolo de reacción ensayado se utilizó araujaína libre y se ensayaron como medios de reacción mezclas orgánico:acuosas 85:15 (Hou *et al.*, 2005). Los solventes orgánicos ensayados fueron ACN y EtOH, y como solvente acuoso se utilizó buffer ácido borico-borato de sodio 0,1M pH 8,0. En este ensayo, se agregó trietilamina (TEA) como agente neutralizante. Se realizaron los correspondientes blancos de enzima y de sustrato, según se detalla en la tabla 3.2. En todos los casos, el volumen final fue 2 ml. La reacción se efectuó a temperatura ambiente y con agitación magnética (300 rpm). Se tomaron muestras (100 µl) a diferentes tiempos (30 minutos, y 1, 2, 3, 4, 5 y 6 horas). La reacción se detuvo con el agregado de 20 µl de ácido acético glacial y 380 µl de MeOH.

Tabla 3.2: Concentraciones finales de los reactivos utilizados para la síntesis enzimática realizada siguiendo el segundo protocolo de reacción.

Medio	Reactivo	Mezcla de reacción	Blanco de enzima	Blanco de sustratos
Solvente orgánico (1,7ml)	Bz-Arg-OEt	0,1 mmol	0,1 mmol	-
	glicinol	0,5 mmol	0,5 mmol	-
	TEA	70 µl	70 µl	70 µl
Buffer (0,3ml)	DTT	10 mg	10 mg	10 mg
	araujaína libre	20 mg	-	20 mg

3.1.6.3. Condiciones de análisis por TLC y HPLC

El progreso de la reacción efectuada bajo ambos protocolos se monitoreó por cromatografía en capa fina (TLC) utilizando cromatofolios de sílica gel GF₂₅₄ (Merck) como fase estacionaria y una mezcla n-butanol:ácido acético:agua (4:1:1) como fase móvil (Hou *et al.*, 2005). El revelador utilizado fue una solución de ácido fosfomolibdico 5% (p/v) en etanol. Las condiciones del análisis se resumen en la tabla 3.3.

Las muestras también fueron analizadas por RP-HPLC en un equipo Äkta Purifier 10 (GE Healthcare) utilizando como fase estacionaria una columna C18 (Phenomenex Nucleosil 5, 150 mm×4,60 mm, 5 µm) y como fase móvil un gradiente de solventes H₂O 0,1% TFA y ACN:H₂O (80:20) 0,1% TFA. Las condiciones del análisis se detallan en la tabla 3.4.

Se analizaron las muestras de reacción, los blancos correspondientes y soluciones patrón del sustrato (Bz-Arg-OEt, 30 mM) y de su correspondiente producto de hidrólisis (Bz-Arg-OH). Éste último fue obtenido por hidrólisis efectuada por papaína (Fluka) de Bz-Arg-OEt 30 mM a 37°C en buffer fosfatos 0,1 M pH 7,0 durante 30 minutos. La enzima fue precipitada con metanol y separada por centrifugación. La Bz-Arg-OH fue recuperada por evaporación del solvente y solubilizada en metanol.

Tabla 3.3: Condiciones de corrida de las cromatografías en capa fina (TLC)

Fase estacionaria	placas de sílica gel GF254 (4 cm×5 cm)
Fase móvil	mezcla n-butanol:ácido acético:agua (4:1:1)
Revelador	ácido fosfomolibdico
Volumen sembrado	5 µl

Tabla 3.4: Condiciones de corrida de la cromatografía líquida de alta resolución en fase reversa (RP-HPLC)

Fase estacionaria	C18 (Phenomenex Nucleosil 5, 150 mm×4,60 mm, 5 μm)
Volumen de columna (VC)	2,5 ml
Fase móvil	Solvente A: H ₂ O 0,1% TFA
	Solvente B: ACN: H ₂ O (80:20) 0,1% TFA
B % inicial	10 %
B % final	20 %
Longitud del gradiente	10 VC
Flujo	1 ml/min
Longitud de onda (nm)	254 y 215
Volumen de inyección	10 μl

3.1.7. Síntesis enzimática de Bz-Arg-Gly utilizando glicina como nucleófilo. Seguimiento por TLC y RP-HPLC

Para la realización de esta reacción (Figura 3.2) se utilizó el protocolo descrito en la sección 3.1.6.2., al cual se le realizaron las siguientes modificaciones: el nucleófilo utilizado fue glicina (Invitrogen), solo se ensayo ACN como solvente orgánico, se modificó la cantidad de trietilamina tal que su concentración fue equimolar a la del Bz-Arg-OEt y no se utilizó DTT. En la tabla 3.5. se detallan las concentraciones finales de cada reactivo en la mezcla de reacción.

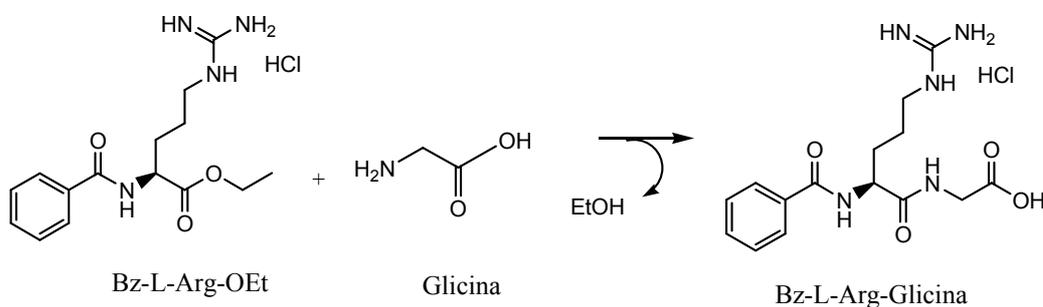


Figura 3.2: Síntesis de Bz-Arg-Gly.

Tabla 3.5: Concentraciones finales de los reactivos utilizados para la síntesis enzimática realizada empleando glicina como nucleófilo.

Medio	Reactivo	Mezcla de reacción	Blanco de enzima	Blanco de sustratos
ACN (1,7ml)	Bz-Arg-OEt	0,1 mmol	0,1 mmol	-
	Glicina	0,5 mmol	0,5 mmol	-
	TEA	0,1 mmol	0,1 mmol	0,1 mmol
Buffer (0,3ml)	araujaína libre	20 mg	-	20 mg

Nuevamente, la reacción fue monitoreada por TLC y RP-HPLC según lo descrito en las tablas 3.3 y 3.4 (sección 3.1.6.3.).

3.1.8. Síntesis enzimática de derivados de Bz-Arg-Ala-OMe utilizando Ala-OMe como nucleófilo. Seguimiento por TLC y RP-HPLC

Para la realización de esta reacción (Figura 3.3) se utilizó el protocolo descrito en la sección 3.1.6.2., al cual se le realizaron las siguientes modificaciones: el nucleófilo utilizado fue Ala-OMe.HCl (Sigma); modificó la cantidad de TEA tal que su concentración final permitiera la neutralización de ambos sustratos. Se ensayaron como solventes orgánicos ACN y EtOH. En la tabla 3.6. se detallan las concentraciones finales de cada reactivo en la mezcla de reacción.

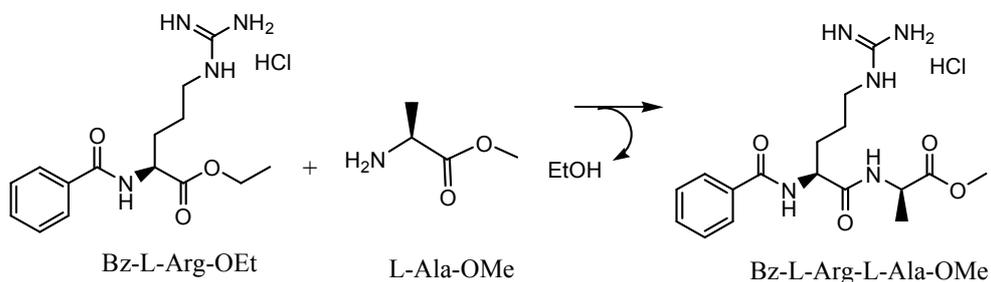


Figura 3.3: Síntesis de Bz-Arg-Ala-OMe.

Tabla 3.6: Concentraciones finales de los reactivos utilizados para la síntesis enzimática realizada empleando Ala-OMe como nucleófilo.

Medio	Reactivo	Mezcla de reacción	Blanco de enzima	Blanco de sustratos
Solvente orgánico (1,7ml)	Bz-Arg-OEt	0,1 mmol	0,1 mmol	-
	Ala-OMe	0,5 mmol	0,5 mmol	-
	Trietilamina	0,6 mmol	0,6 mmol	0,6 mmol
Buffer (0,3ml)	DTT	10 mg	10 mg	10 mg
	Araujiaína libre	20 mg	-	20 mg

Nuevamente, la reacción fue monitoreada por TLC y RP-HPLC según lo descrito en las tablas 3.3 y 3.4 (sección 3.1.6.3.).

3.1.9. Análisis de los productos obtenidos mediante HPLC-ESI MS

El análisis de los productos obtenidos en la reacción de síntesis utilizando Ala-OMe como nucleófilo se realizó mediante HPLC acoplado a un detector de espectrometría de masas electrospray. Las características del equipo y las condiciones de análisis se detallan en la tabla 3.7.

Tabla 3.7: Características del equipo y las condiciones de análisis.

HPLC	Agilent 1100 LC/MSD
Detector	UV-Vis del LC (DAD)
Fase Estacionaria	C18 (Phenomenex Nucleosil 5, 150 mm×4,60 mm, 5 µm)
Volumen de columna (VC)	2,5 ml
Fase móvil	Solvente A: H ₂ O 0,1% TFA Solvente B: ACN: H ₂ O (80:20) 0,1% TFA
B % inicial	10 %
B % final	20 %
Longitud de onda (nm)	254 y 215
Volumen de inyección	20 µl
Espectrómetro de masas	Agilent (G 1956A VL; ES-MS; APCI-MS)
Velocidad de flujo fase móvil LC	0,5 ml/min
Flujo del gas secante	13 l/min
Temperatura del nebulizador	60°C
Temperatura del gas secante	350°C.
Voltaje del capilar de muestreo	4000 V

3.2. Aplicación de la metodología 3D-QSAR para la obtención de modelos cualitativos de predicción de actividad sobre receptores IN de los derivados Bz-Arg.

3.2.1. Generación del conjunto de entrenamiento.

3.2.1.1. Compuestos utilizados

Se utilizaron 40 estructuras (vease 8.2.1.) de RGD mimetics tomadas de bibliografía (Gentilucci *et al.* 2007; Breslin *et al.* 2004; Bourguet *et al.*, 2003; Coleman *et al.* 2002; Lange *et al.* 2002; Zechel *et al.* 2002). El parámetro de actividad biológica utilizado fue la concentración inhibitoria media (IC_{50}) que dichos compuestos presentan para las integrinas sobre las que fueron ensayadas. El valor de corte adoptado para diferenciar compuestos activos de inactivos fue de 500 nM. De esta forma, los compuestos fueron clasificados en dos categorías:

~ activos: si $IC_{50} < 500$ nM

~ inactivos: si $IC_{50} \geq 500$ nM

Se utilizó el programa HyperChem6.03® para representar tridimensionalmente las estructuras de los compuestos utilizados.

3.2.2. Cálculo de los descriptores

3.2.2.1. Optimización conformacional

Para realizar el modelado empleando descriptores 3D es necesario partir de una representación molecular correspondiente a una conformación probable. En consecuencia, y a fin de optimizar la conformación de cada una de las estructuras del conjunto de entrenamiento, se realizaron dos etapas secuenciales de optimización, utilizando el programa HyperChem6.03®. En primera instancia, se empleo un método de cálculo basado en mecánica molecular (método MM+). Dado que el conjunto de compuestos utilizados

para el modelado presentan numerosos enlaces con libre rotación, resulta inviable la realización de un estudio conformacional sistemático (Chang *et al.*, 1989). Por lo tanto, se utilizó el módulo Conformational Search del programa Hyperchem para realizar una exploración estocástica –tipo Montecarlo– del espacio conformacional, utilizando el método Usage Directed (Chang *et al.*, 1989). Esta rutina genera, a partir de un conformero inicial, un nuevo conformero, variando aleatoriamente un número al azar de ángulos dihedros (en nuestro caso, entre 1 y 8 ángulos). Luego, esta estructura es optimizada para llegar al conformero de mínima energía más próximo en la hipersuperficie de energía. Este conformero es sometido a su vez a variación estocástica de sus ángulos dihedros, y el conformero resultante es optimizado nuevamente. Este ciclo se repite y de esta manera, se va generando una lista de los conformeros de mínima energía encontrados. El método usage-directed muestrea uniformemente la lista de conformeros de mínima energía generada para encontrar nuevos conformeros; para ello se va utilizando como estructura de partida para aplicar la variación estocástica el conformero de mínima energía que menos veces se ha utilizado con anterioridad para este fin. Los cinco conformeros de menor energía así obtenidos fueron luego refinados mediante un método semiempírico de cálculo. Los métodos y parámetros utilizados en cada etapa de optimización se detallan en la tabla 3.8. Finalmente, se elige el conformero de menor energía de entre los cinco refinados para calcular los valores que asumen los descriptores moleculares 3D del programa Dragon para cada compuesto.

Tabla 3.8: Parámetros de cálculo utilizados en la optimización conformacional.

Tipo de cálculo	Método	Algoritmo de cálculo	Gradiente RMS (kcal/Å mol)
Mecanica Molecular (MM+)	Usage Directed	Polak-Ribiere (gradientes conjugados)	0,1
Semiempirico (PM3)	-	Polak-Ribiere (gradientes conjugados)	0,05

3.2.2.2. Cálculo del valor de los D descriptores

Para la construcción de los modelos, se emplearon descriptores 0, 1, 2 y 3D, calculados mediante el programa Dragon4.0®.

Empleando el conjunto de entrenamiento conformado por 40 estructuras optimizadas (22 activas y 18 inactivas), y excluyendo aquellos descriptores cuyos valores resultaban constantes o cuasi-constantes entre estas estructuras, se obtuvieron 20 bloques de descriptores (1612 descriptores totales); agrupados según la definición y el método de cálculo de los mismos (por ejemplo, el bloque correspondiente a índices de conectividad contiene un conjunto de descriptores calculados mediante operaciones algebraicas sobre la matriz de conectividad del grafo) (Talevi, 2007).

3.2.3. Modelado: Selección de un subconjunto de d descriptores

Se realizaron 36 combinaciones al azar de los 20 bloques de descriptores calculados en el punto anterior. Cada combinación incluía al menos un bloque de descriptores 3D y el número total de descriptores fue de entre 150 y 200. Se requiere incluir descriptores 3D ya que durante la búsqueda bibliográfica se observó que existen isómeros ópticos análogos de la secuencia RGD con valores de actividad por encima y por debajo del valor de corte definido; en otras palabras, la búsqueda bibliográfica indicó que la configuración de los isómeros ópticos influye en la actividad sobre los receptores de integrinas, y sólo los descriptores dependientes de la conformación permiten discriminar entre isómeros ópticos (Gentilucci *et al.* 2007).

Utilizando los módulos regresión lineal múltiple y análisis lineal discriminante del programa Statistica7.0®, y la técnica Forward Stepwise (vease 1.2.5.3.1.), se evaluaron las 36 combinaciones de descriptores. Debido a que nuestro conjunto de entrenamiento está conformado por 40 compuestos, y considerando la regla empírica que vincula la cantidad de compuestos del conjunto de entrenamiento con el número de descriptores del modelo (vease 1.2.5.3.3.), el número de pasos utilizados fue de 3. Por otro lado, el valor de tolerancia β aceptado fue 0,3, a fin de no incluir pares de descriptores redundantes.

Los métodos de cálculo y parámetros utilizados se detallan en la tabla 3.9.

Tabla 3.9: Parámetros de cálculo utilizados en el modelado

Tipo de cálculo	Método	Tolerancia	Número de pasos
Regresión lineal múltiple	Forward stepwise	0,3	3
Análisis discriminante	Forward stepwise	0,3	3

En ambos tipos de análisis, la variable dependiente utilizada fue la variable binaria, df (función discriminante), cuyo valor asignado a los compuestos activos fue de 1, y a los inactivos -1. En la regresión múltiple, todos los descriptores de la combinación de descriptores analizada fueron elegidos como variables independientes y los 3 descriptores que se obtuvieron de este análisis fueron luego utilizados como variables independientes durante el análisis discriminante.

Para elegir la mejor df de entre las funciones generadas se utilizaron los siguientes criterios:

- ✓ observación del % de buenas clasificaciones en el conjunto de entrenamiento (entre el 85 y el 95% de predicciones correctas tanto para los compuestos activos como para los compuestos inactivos)
- ✓ observación del valor del parámetro λ de Wilk (se descartaron los modelos cuyo λ de Wilk $\geq 0,5$)

3.2.4. Validación

Los modelos obtenidos en este trabajo fueron sometidos a validación interna y externa.

3.2.4.1. Validación Externa

El 30% del total de las estructuras disponibles fueron utilizadas para esta validación y se encuentran representadas en la sección 8.2.2.

Los modelos obtenidos fueron luego utilizados para predecir el valor de df de compuestos de IC_{50} conocida.

3.2.4.2. Validación Interna

Se realizó mediante el ensayo de aleatorización de Fisher y la validación cruzada Leave-Group-Out.

Para efectuar el ensayo de aleatorización de Fisher, se realizaron 40 aleatorizaciones de la variable df para cada combinación de bloques de descriptores. La calidad estadística del modelo original se comparó con la de los modelos aleatorizados así obtenidos, mediante la comparación del valor F , el % de buenas clasificaciones y el estadístico λ de Wilk.

Por otro lado, se realizó el ensayo de validación cruzada Leave-Group-Out. Para ello, 5 compuestos elegidos al azar fueron excluidos del conjunto de entrenamiento. Se realizó el análisis discriminante utilizando como variables independientes aquellos descriptores que constituían el modelo original y la nueva regresión obtenida se utilizó para predecir el valor de df de los 5 compuestos excluidos del conjunto de entrenamiento y del conjunto de compuestos utilizados para la validación externa. Este procedimiento se realizó 16 veces para cada modelo.

3.2.5. Análisis del modelo seleccionado

Se construyó un Diagrama de Distribución Farmacológica para los compuestos del conjunto de entrenamiento (Gálvez *et al.*, 1996) del modelo que superó las instancias de

validación. En estos gráficos se muestra la expectativa de hallar un compuesto activo, o inactivo, en función del valor de la función discriminante. A continuación se encuentran las expresiones que permiten calcular ambas expectativas.

$$\text{Expectativa activo} = \frac{\% \text{ de activos}}{\% \text{ de inactivos} + 100}$$

$$\text{Expectativa inactivo} = \frac{\% \text{ de inactivos}}{\% \text{ de activos} + 100}$$

Donde el % de activos y el % de inactivos se definen según:

$$\% \text{ de activos} = \frac{\text{número de activos en el intervalo considerado de la } df}{\text{número de activos totales utilizados para generar la } df} \times 100$$

$$\% \text{ de inactivos} = \frac{\text{número de inactivos en el intervalo considerado de la } df}{\text{número de inactivos totales utilizados para generar la } df} \times 100$$

Por otro lado, se analizaron los valores de los coeficientes de regresión y el significado de los descriptores que constituyen el modelo seleccionado, con el objetivo de hacer una interpretación físico-química de los mismos.

3.2.6. Aplicación del modelo obtenido para la predicción de actividad sobre receptores IN de los derivados Bz-Arg.

El modelo seleccionado luego de ambos procesos de validación, fue utilizado para predecir la actividad de los productos de síntesis de las reacciones ensayadas.

4. Resultados y Discusión

4.1. Síntesis enzimática de derivados de N^α -benzoil arginina (Bz-Arg)

4.1.1. Caracterización de araujiaina

Los ensayos de caracterización de la muestra vegetal fueron realizados según lo descrito en la sección 3.1.3.

4.1.1.1. Determinación de la cantidad de proteína por el método de Bradford

La concentración de proteína de la solución 1,5 mg de araujiaina/ml se determinó a partir de la curva de calibración realizada con BSA y el resultado obtenido se muestra en la tabla 4.1.

Tabla 4.1: Resultado de la determinación de cantidad de proteína por el método de Bradford.

Muestra	Concentración de proteína (mg/ml)
Solución 1,5 mg de solido/ml	0,828

4.1.1.2. Determinación de actividad enzimática

La actividad proteolítica de la muestra vegetal (araujiaina) se determinó usando caseína como sustrato y expresa en Ucas y Ucas/mg de proteína (Tabla 4.2), siendo 1 Uca la cantidad de enzima que produce un incremento de 1 unidad de Absorbancia a 280 nm, en 1 ml de muestra, en 1 minuto de reacción, en las condiciones de ensayo.

Tabla 4.2: Resultado de la determinación de Actividad caseinolítica.

Muestra	Actividad caseinolítica (Ucas)	Actividad específica (Ucas/mg proteína)
Solución 1,5mg de liofilizado/ml	34,04	22,69

4.1.1.3. Ensayos de estabilidad y actividad caseinolítica residual

Con el fin de optimizar la determinación de la actividad caseinolítica residual, se utilizó araujaína en buffer ácido bórico-borato de sodio 0,1 M pH 8,0. Los resultados de los tiempos de reacción ensayados (véase 3.1.3.3.) se encuentran en la tabla 4.3.

Tabla 4.3: Resultado de la determinación de actividad caseinolítica del control a diferentes tiempos de reacción.

Tiempo (min)	A ₂₈₀
2	0,0804
5	0,2999
10	0,6196

Según estos resultados, y debido a que se espera una caída en la actividad de araujaína en las muestras tomadas de medios orgánico:acuoso, se eligió como tiempo de reacción caseinolítica 10 minutos.

Para calcular la actividad caseinolítica residual porcentual se considero como 100% de actividad la presentada por el control (araujaína en buffer, tiempo cero). Los resultados obtenidos se muestran en la figura 4.1.

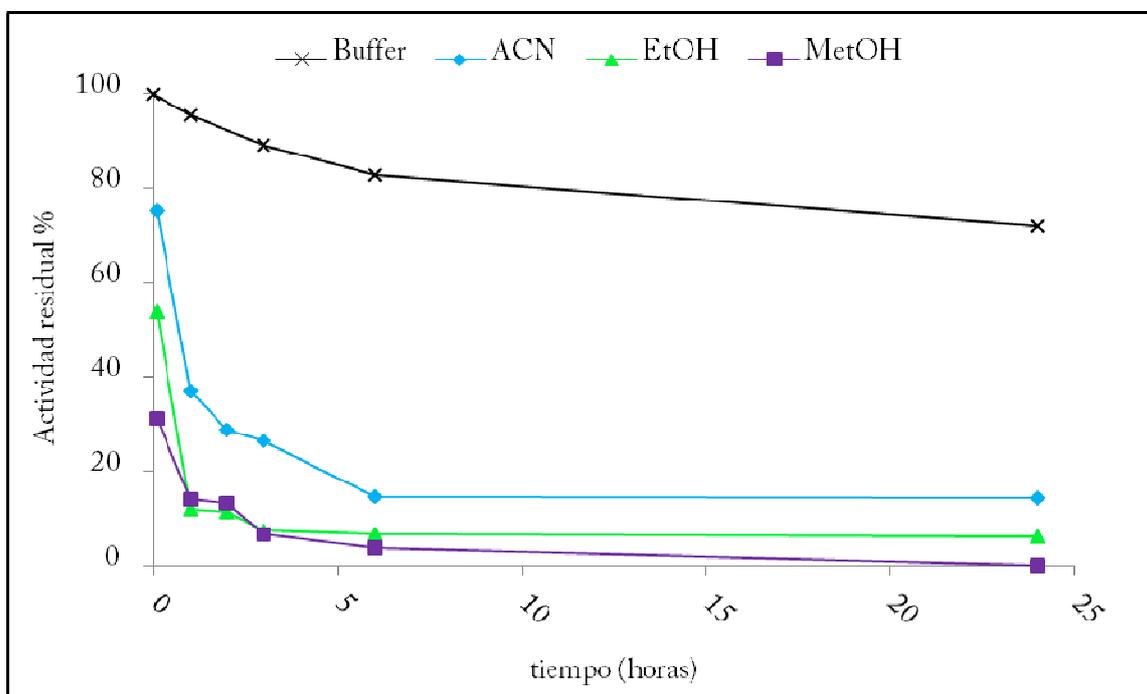


Figura 4.1: Actividad caseinolítica residual porcentual de preparaciones enzimáticas en las diferentes mezclas orgánico:acuosas.

Como esperabamos, no hubo gran pérdida de actividad enzimática en el medio utilizado como control (buffer). Luego de 24 horas de incubación a temperatura ambiente la caída de actividad fue menor al 30%.

Las curvas obtenidas revelan que araujaína se comporta en forma semejante en las diferentes mezclas orgánico:acuosas. Puede observarse que la mayor caída de actividad se produce durante la primera hora de contacto, luego de lo cual hay un breve periodo de cuasi-estabilización, seguido de una nueva caída de actividad. En todos los casos, la pérdida de actividad prácticamente se detuvo a las 6 horas.

Medio MeOH:buffer: el contacto de la enzima durante apenas 5 minutos es suficiente para causar una caída de la actividad de araujaína de casi un 70%. Transcurridas 6 horas, la actividad es menor al 10% y a las 24 horas es menor al 1%.

Aunque se ha encontrado varios reportes del uso de medios conteniendo mediano y alto contenido de MeOH en síntesis empleando proteasas (Hou *et al.*, 2005), es evidente

que araujiáina pierde completamente su capacidad de catálisis. Es probable que la toxicidad del MeOH para la mayoría de las peptidasas se deba al desplazamiento del agua de hidratación que las estabiliza.

Medio EtOH:buffer: si bien la caída de actividad es tan pronunciada como la observada en MetOH, probablemente debido a las mismas razones, el contacto inicial con el medio solo causó un 45% de pérdida de actividad. Luego de 6 horas araujiáina aún conservó cerca del 10% de su actividad inicial.

Se ha encontrado en bibliografía que papaína pierde el 20% de actividad luego de 20 minutos de incubarla en presencia de 90% de EtOH (Szabó *et al.*, 2006). En contraparte, Chen *et al.* (1992) informaron que la alcalasa que su grupo utilizó retuvo la mitad de su actividad original después de 5 días en EtOH.

Medio ACN:buffer: de los medios ensayados, éste mostro ser el menos agresivo para araujiáina. Incluso luego de 24 horas de incubación, araujiáina retuvo un 15% de su actividad inicial.

Hay múltiples reportes del empleo de ACN como solvente orgánico empleado para la síntesis de péptidos y otros derivados peptídicos catalizada por enzimas proteolíticas, no sólo por permitir que los biocatalizadores manifiesten su actividad en mayor o menor medida, sino porque tanto sustratos como productos de las reacciones son solubles en este solvente (Morcelle *et al.*, 2008; Hou *et al.*, 2005).

En base a los resultados de este ensayo, se decidió utilizar la mezcla ACN:buffer como principal medio de reacción. Por otro lado, considerando que el EtOH es un disolvente orgánico miscible con el agua y conveniente para la síntesis de péptidos que contienen aminoácidos hidrofílicos, no muestra problemas de toxicidad, su costo es bajo y puede ser fácilmente eliminado del sistema de reacción, también fue ensayado como medio de síntesis. En ambos casos, el tiempo máximo de reacción fue 6 horas.

4.1.2. Determinación de la actividad amidolítica de araujaína inmovilizada

Las unidades enzimáticas del extracto proteolítico inmovilizado fueron determinadas usando BAPNA como sustrato. La actividad fue expresada en Unidades Internacionales (UI) y en UI/mg de inmovilizado, siendo una UI la cantidad de enzima que produce 1 μmol de producto (*p*-nitroanilina) en un minuto de reacción, en las condiciones de ensayo (véase 3.1.5.). Los resultados se encuentran en la tabla 4.4.

Tabla 4.4: Resultado de la determinación de actividad amidolítica de araujaína inmovilizada.

UI	UI/mg araujaína inmovilizada
$3,316 \times 10^{-2}$	$1,105 \times 10^{-3}$

4.1.3. Síntesis enzimática de derivados de Bz-Arg-Glyol utilizando glicinol como nucleófilo. Seguimiento por TLC y RP-HPLC.

Esta síntesis se realizó siguiendo dos protocolos alternativos.

4.1.3.1. Primer protocolo de reacción: enzima inmovilizada y medio de bajo contenido acuoso

Las figuras 4.2-4.6 muestran los resultados obtenidos empleando como medio de reacción ACN 0,25% (v/v) de agua y araujaína inmovilizada en poliamida como biocatalizador. Las condiciones de análisis utilizadas fueron detalladas en las tablas 3.3 y 3.4. (sección 3.1.6.3.)

TLC



Figura 4.2: Análisis por TLC de la reacción Bz-Arg-OEt con Glicinol en ACN 0,25% H₂O. Punto de siembra 1: patrón de sustrato (Bz-Arg-OEt). Punto de siembra 2: patrón de sustrato hidrolizado (Bz-Arg-OH). Punto de siembra 3: muestra de reacción. Punto de siembra 4: blanco de enzima.

La corrida cromatográfica correspondiente al primer punto de siembra muestra una mancha principal ($R_f=0,5$) que se correspondería con el sustrato (Bz-Arg-OEt). Una mancha con R_f similar puede observarse en la corrida de la muestra de reacción (punto de siembra 3) y en la de la mezcla de sustratos (punto de siembra 4). Esta misma corrida también presenta una mancha pequeña, cerca del punto de siembra ($R_f=0,13$) que podría tratarse de impurezas presentes, y que también se observa en el resto de las corridas de esta TLC.

Como se esperaba, el desarrollo de la muestra sembrada en cuarto lugar, que corresponde a la mezcla de sustratos, muestra el mismo perfil que el observado para el patrón de sustrato.

Respecto a la corrida del segundo punto de siembra, donde se sembró el patrón de sustrato hidrolizado (Bz-Arg-OH), pueden observarse tres manchas: la de menor R_f (0,13), que correspondería a las mismas impurezas observadas en las demás corridas, la de R_f intermedio ($R_f=0,38$) que correspondería al sustrato hidrolizado, y la de mayor R_f (0,56) que correspondería a sustrato no hidrolizado.

El desarrollo de la muestra de reacción presenta una mancha de R_f similar a la del sustrato ($R_f=0,44$) y una mancha que correspondería a sustrato hidrolizado ($R_f=0,33$). Puede observarse también una mancha intermedia ($R_f=0,42$) que, considerando las

polaridades del sustrato, el producto de hidrólisis y el producto de síntesis, se esperaba correspondiera a Bz-Arg-Glyol. Los resultados del análisis mediante RP-HPLC se encuentra en las Figuras 4.3-4.6.

RP-HPLC

Los cromatogramas obtenidos para soluciones patron del sustrato y de su producto de hidrólisis se muestran superpuestos en la figura 4.3. En el anexo 8.1.1. se encuentran estos mismos resultados, en gráficas individuales.

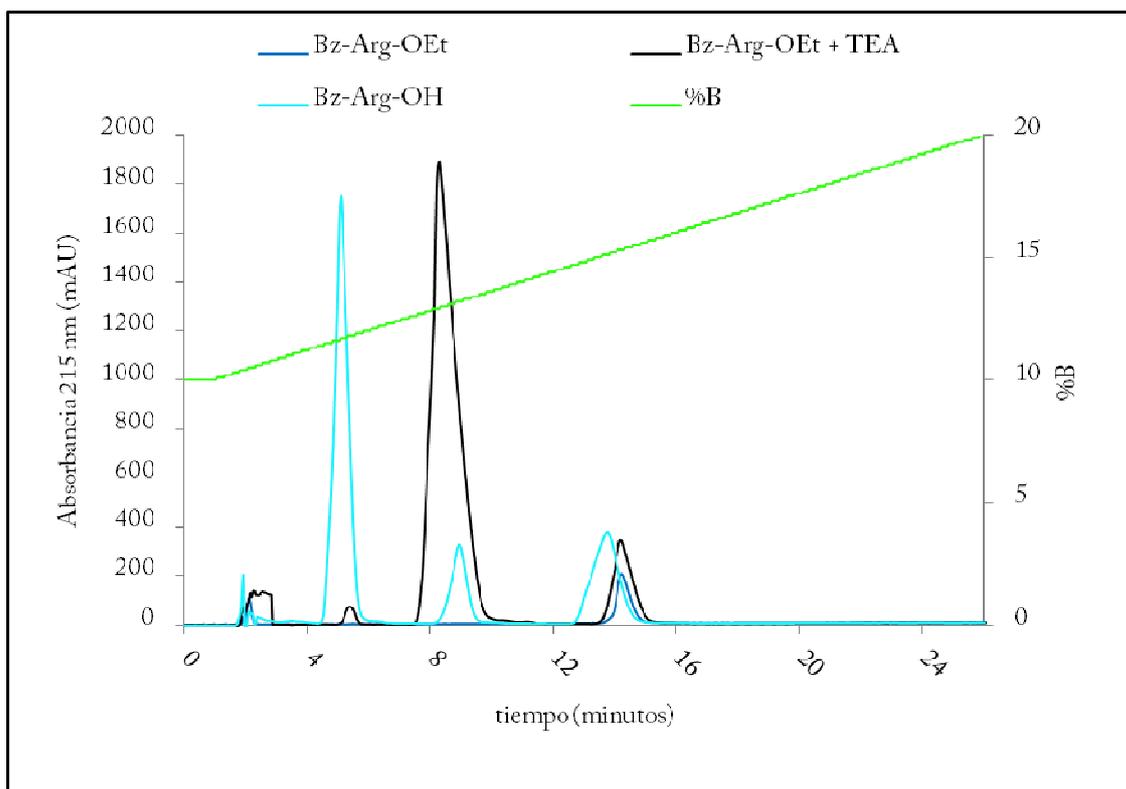


Figura 4.3: Cromatogramas correspondientes a los patrones de sustrato (Bz-Arg-OEt y Bz-Arg-OEt conteniendo TEA en cantidad equimolar) y de sustrato hidrolizado (Bz-Arg-OH).

El patrón de sustrato muestra un pico principal cuyo tiempo de retención es de 14,2 minutos en estas condiciones de análisis.

El patrón de sustrato presenta un pico principal, cuyo tiempo de retención es 5,10 minutos que correspondería a la Bz-Arg-OH (el tiempo de retención de este ácido se espera sea mucho menor al tiempo de retención de su correspondiente ester etílico). También puede observarse un pico cercano a los 14 minutos que se correspondería con sustrato no hidrolizado. Además puede apreciarse otro pico, de tiempo de retención intermedio, cuya presencia no se ha podido explicar, pero que también puede observarse en el cromatograma obtenido para el patrón de sustrato que contiene trietilamina (TEA) en cantidad equimolar. Presumiblemente, este pico podría tratarse del producto de una reacción secundaria del sustrato o incluso de su producto de hidrólisis.

En el anexo 8.1.2.1. se encuentran todos los cromatogramas obtenidos en gráficas individuales. En esta sección (figura 4.4) se muestran superpuestos los perfiles obtenidos para una muestra de reacción y del blanco de enzima tomadas luego de 3 horas de reacción.

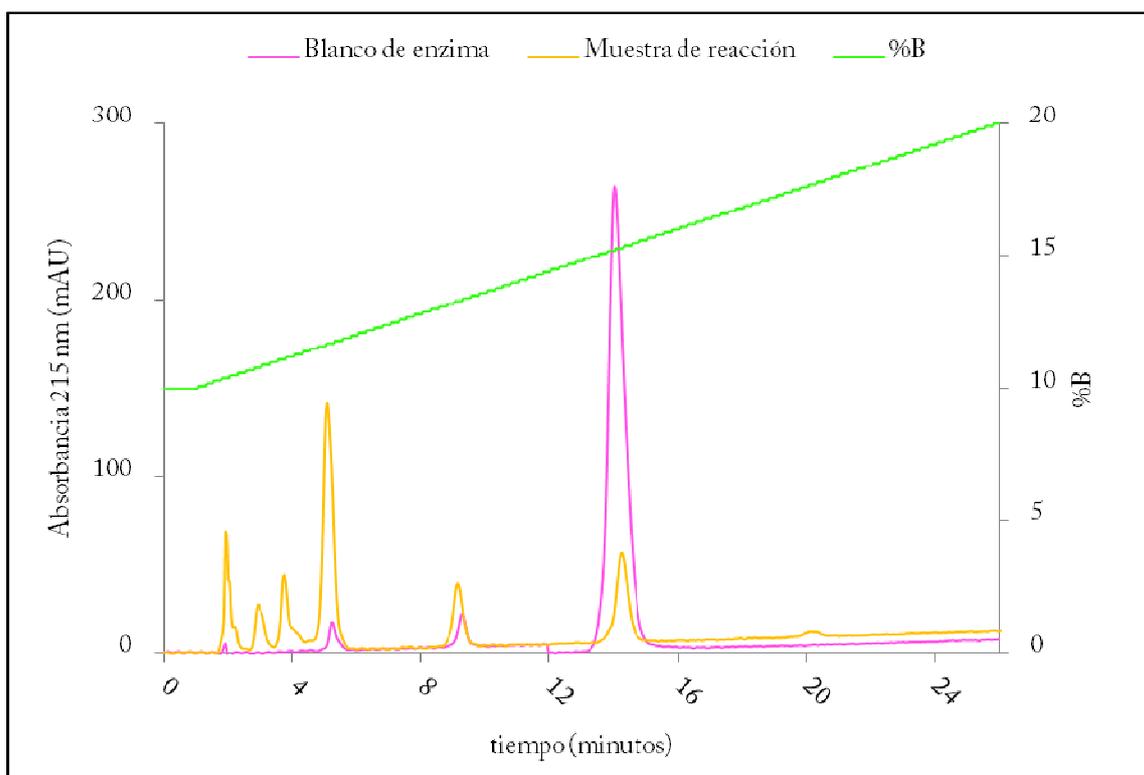


Figura 4.4: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en ACN 0,25% H₂O. Cromatogramas correspondientes al blanco de enzima y a una muestra de reacción tomada luego de 3 horas de incubación.

Los cromatogramas obtenidos muestran el mismo perfil. Pueden observarse los picos correspondientes a la Bz-Arg-OEt (t_R 14 minutos) y a la Bz-Arg-OH (t_R 5,10 minutos). La presencia de una importante cantidad del producto de hidrólisis estaría indicando que la enzima se encuentra activa.

Por otro lado puede observarse un pico de 9 minutos de tiempo de retención que coincide con el observado en el patrón de Bz-Arg-OH y en el patrón de Bz-Arg-OEt conteniendo TEA. La presencia de un compuesto de tiempo de retención intermedio entre el observado por el sustrato y el observado para el sustrato hidrolizado podría explicar también la presencia de la mancha de R_f intermedio observada en la TLC (corrida correspondiente al punto de siembra 3). Sin embargo se descarta que se trate del producto de síntesis debido a que se encuentra presente en dichos patrones y en el blanco de enzima.

Por otro lado, en los tres cromatogramas que corresponden a las muestras de reacción tomadas luego de 1, 3 y 6 horas de reacción (Figuras 4.4 y 4.5) aparecen una serie de picos de bajo tiempo de retención (menores a 5 minutos) que no están presentes en el blanco de enzima. Los mismos podrían tratarse de productos de reacciones secundarias, no esperadas, o de impurezas que podrían provenir de la enzima inmovilizada. Para descartar esta última hipótesis, se analizó el blanco de sustratos (Figura 4.6).

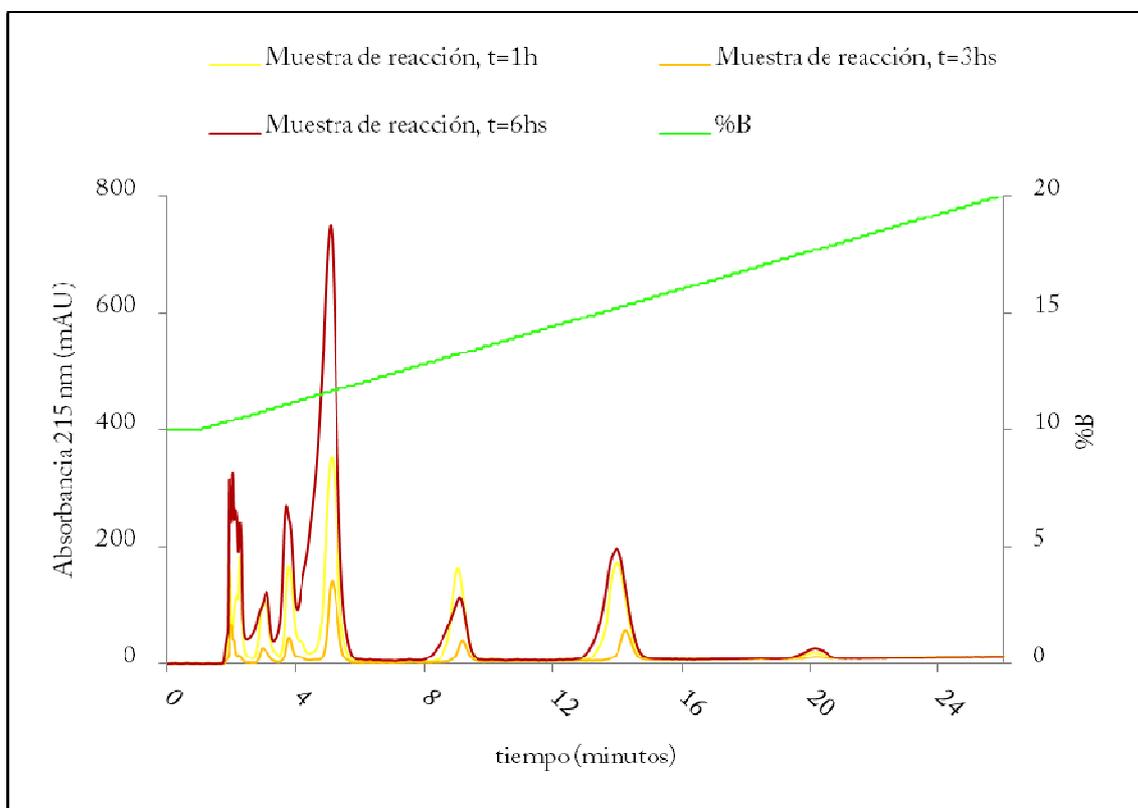


Figura 4.5: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en ACN 0,25% H₂O. Cromatogramas correspondientes tres muestra de reacción, tomada luego de 1, 3 y 6 horas de incubación.

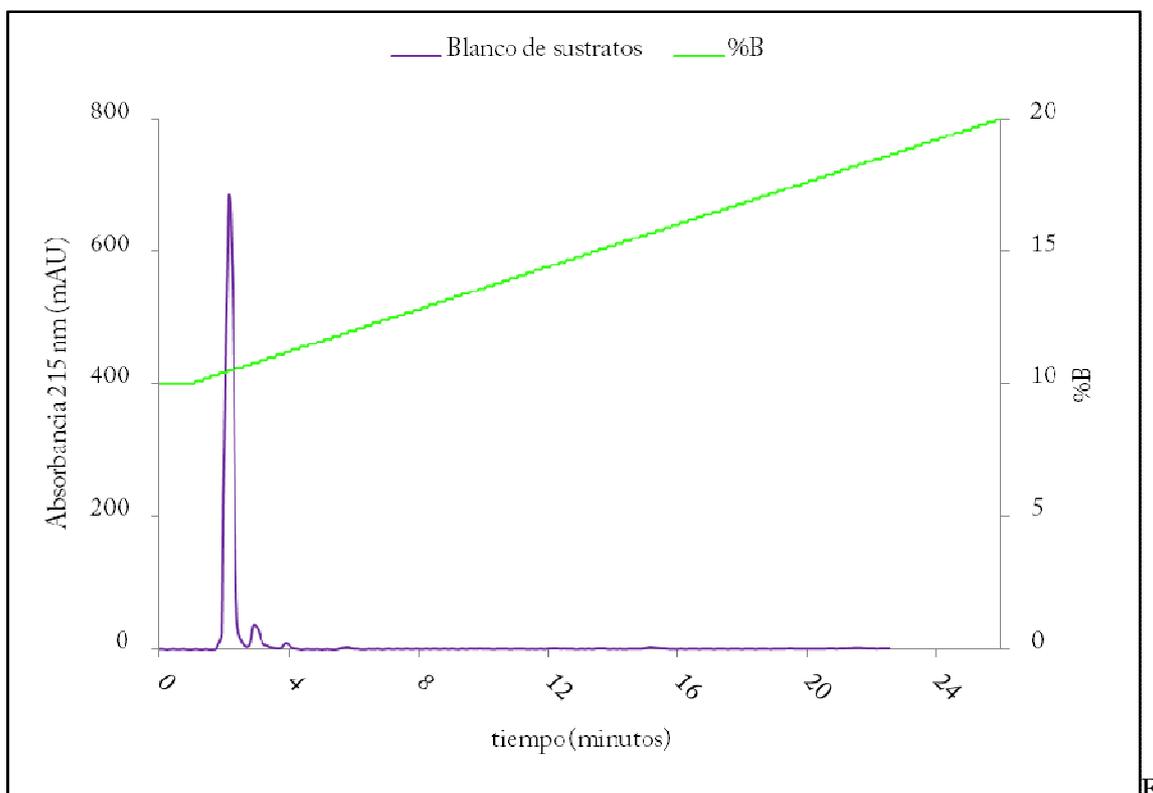


Figura 4.6: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en ACN 0,25% H₂O. Cromatograma correspondiente al blanco de sustratos de la reacción.

En el cromatograma obtenido para el blanco de sustratos (Figura 4.6.) se observa un pico relevante que coincide con el frente de solvente y dos pequeños picos. Este perfil no permite explicar la procedencia de los picos pequeños y de bajo tiempo de retención observados en los cromatogramas anteriores (Fig 4.4 y 4.5), y por ello no se descarta que los picos de tiempos de retención 3 minutos y 3,7 minutos observados en dichos graficos sean resultado de reacciones secundarias, entre ellas, dimerización, descomposición o hidrólisis de algún otro grupo químico del sustrato.

A partir del análisis efectuado, podría concluirse que ninguno de los productos formados correspondería con el producto de síntesis.

4.1.3.2. Segundo protocolo de reacción: enzima libre y medio de mediano contenido acuoso

En este ensayo de síntesis se empleó araujiína libre y dos medios de reacción alternativos. Los resultados obtenidos para la mezcla orgánico:acuosa (85:15) conteniendo ACN se discuten en primer lugar (Figuras 4.7-4.9), y en segundo lugar se analizan los obtenidos para la mezcla que contiene EtOH como solvente orgánico (Figuras 4.10 y 4.11).

Medio ACN:H₂O (85:15)

TLC

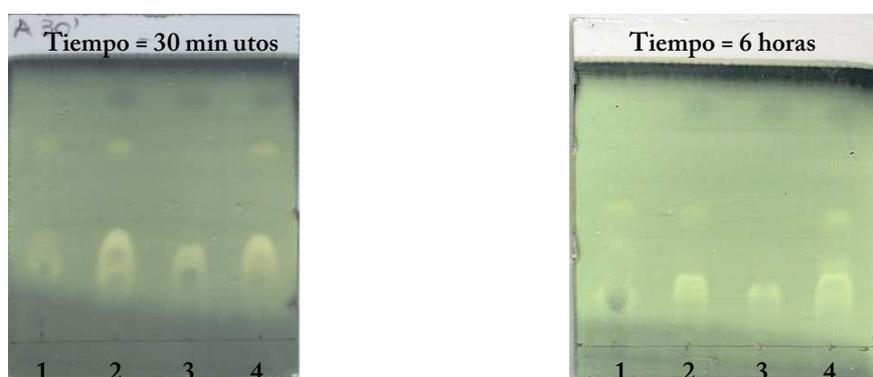


Figura 4.7: Análisis por TLC de la reacción Bz-Arg-OEt con glicinol en la mezcla ACN:H₂O (85:15). Puntos de siembra 1: patron de sustrato (Bz-Arg-OEt). Puntos de siembra 2: blanco de enzima. Puntos de siembra 3: blanco de sustratos. Puntos de siembra 4: muestras de reacción correspondientes.

Las corridas cromatográficas de los blancos y la mezcla de reacción tomadas a 30 minutos revelan perfiles similares. En ambos casos, el patron de sustrato, el blanco de enzima y la muestra de reacción presentan las mismas manchas. Las manchas de menor R_f (0,32) y las de mayor R_f (0,89) corresponderían a impurezas contenidas en las muestras sembradas. También se aprecia la mancha correspondiente al sustrato, que en estas corridas presenta un R_f algo mayor (0,70).

Por otro lado, la muestra de reacción tomada a 6 horas revela una mancha de R_f intermedio (0,42) que no se observa en el patron de sustratos ni en los blancos. Al tener menor R_f que el patron de sustratos, el origen de la misma podría ser la presencia del producto de síntesis, sin embargo el producto de hidrólisis también tiene un R_f similar

(0,38). Para descartar esta última hipótesis, se realizó una nueva TLC (Figura 4.8), en la cual en el punto de siembra número 2 fueron sembrados 5 µl del patrón de Bz-Arg-OH.

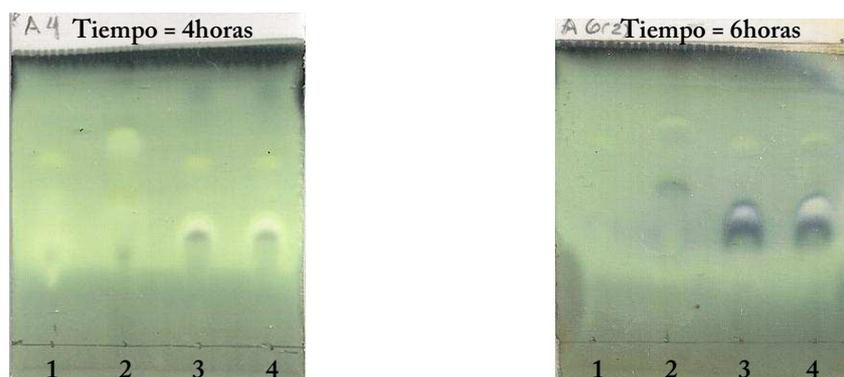


Figura 4.8: Análisis por TLC de la reacción Bz-Arg-OEt con glicinol en la mezcla ACN:H₂O (85:15). Puntos de siembra 1: Patrón de sustrato (Bz-Arg-OEt). Puntos de siembra 2: Patrón de producto de hidrólisis (Bz-Arg-OH). Puntos de siembra 3: Blanco de sustratos. Puntos de siembra 4: Muestras de reacción correspondientes.

Estas nuevas placas demostraron que el origen de dicha mancha era la presencia de producto de hidrólisis en el tubo de reacción.

RP-HPLC

Los análisis efectuados por RP-HPLC (Figura 4.9) confirmaron los resultados evidenciados por TLC.

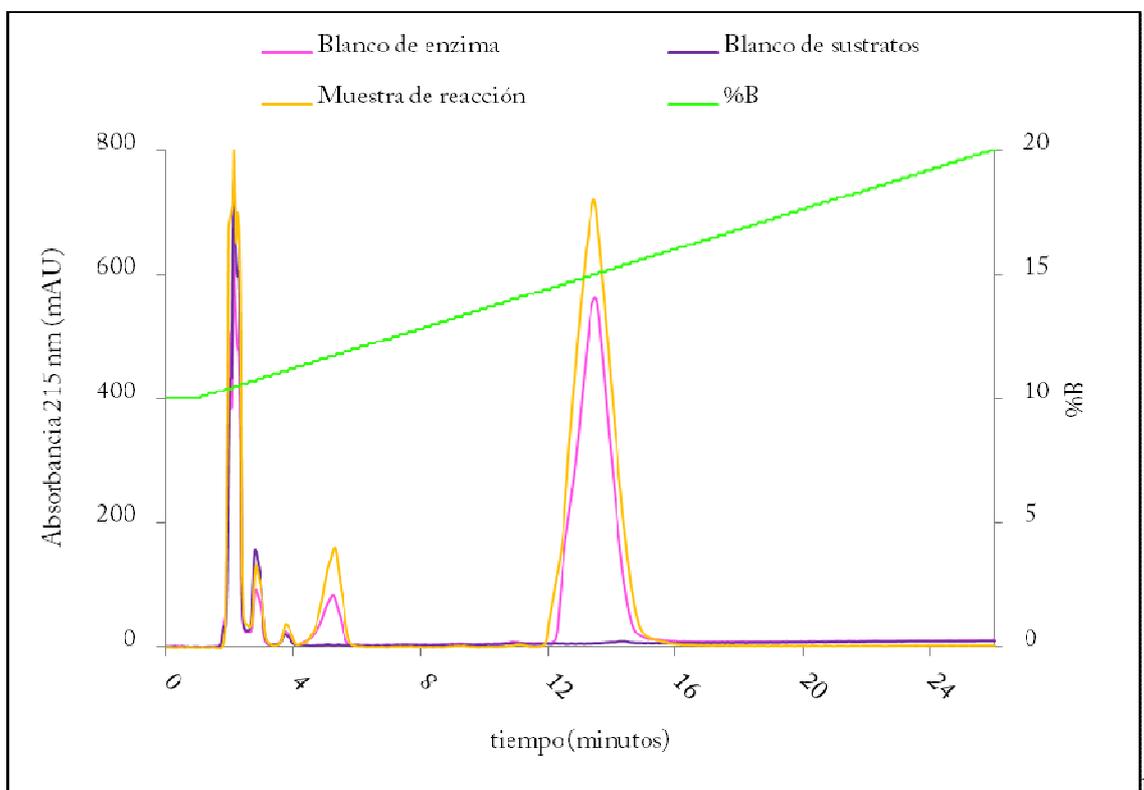


figura 4.9: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en la mezcla ACN:H₂O (85:15). Cromatogramas correspondientes al blanco de enzima, blanco de sustratos y muestra de reacción, luego de 3 horas de reacción.

Los cromatogramas correspondientes al blanco de enzima y a la muestra tomada luego de 3 horas de reacción son prácticamente coincidentes. En ambos casos puede observarse el pico correspondiente al sustrato (t_R 14 minutos) y el correspondiente al sustrato hidrolizado (t_R 5 minutos). Por otro lado, en los tres cromatogramas pueden observarse picos de tiempos de retención menores a 4 minutos que podrían corresponderse con las mismas impurezas observadas cerca del punto de siembra en las TLC.

En el anexo 8.1.2.2. se encuentran todas las gráficas individuales obtenidas durante este análisis, para ambos medios de reacción ensayados.

Medio EtOH:H₂O (85:15)

TLC

Luego de 30 minutos y 6 horas, la muestra de reacción mostró las mismas manchas que el patrón de sustratos y el blanco de enzima (Figura 4.10)

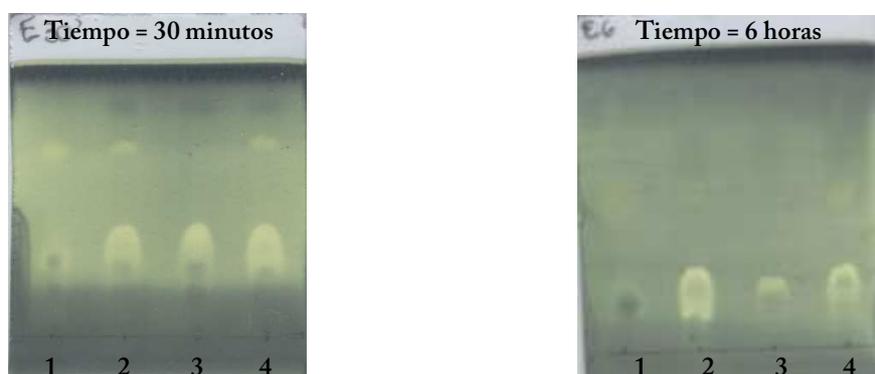


Figura 4.10: Análisis por TLC de la reacción Bz-Arg-OEt con glicinol en la mezcla EtOH:H₂O (85:15). Puntos de siembra 1: Patrón de sustrato (Bz-Arg-OEt). Puntos de siembra 2: Blanco de enzima. Puntos de siembra 3: Blanco de sustratos. Puntos de siembra 4: Muestras de reacción correspondientes.

Los resultados del análisis por HPLC de la muestra de reacción y los blancos correspondientes se muestran en la figura 4.11.

RP-HPLC

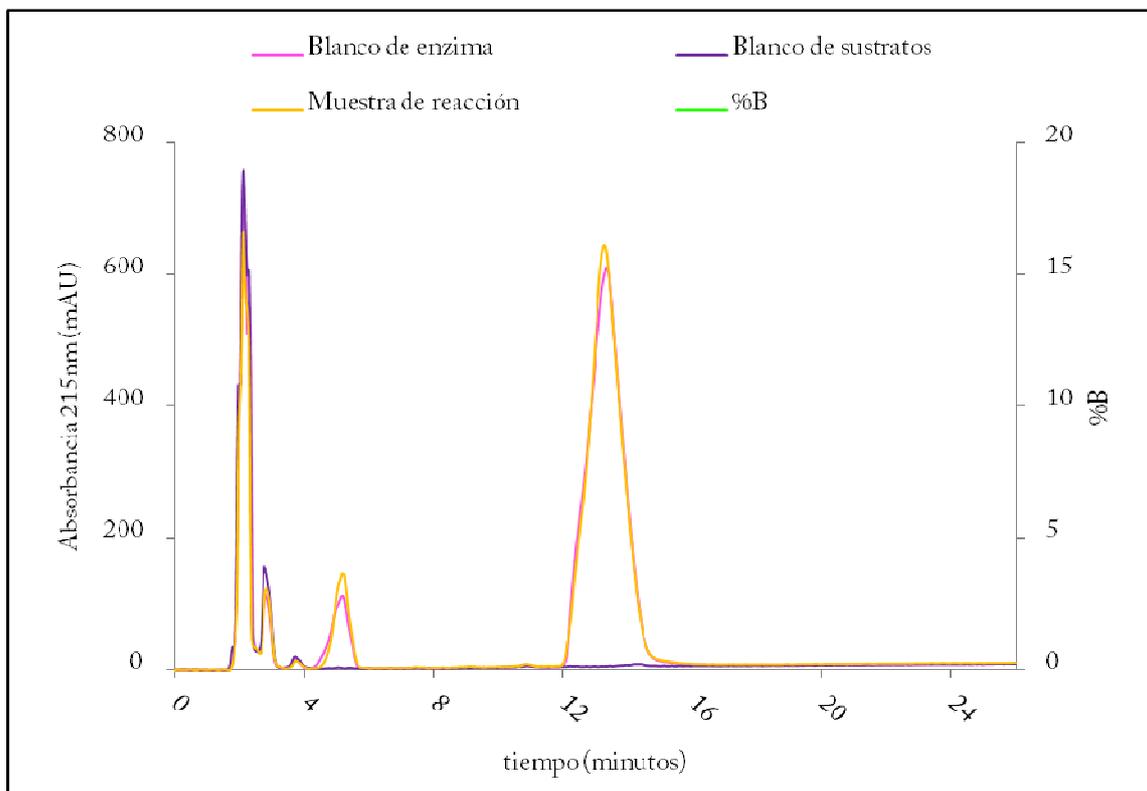


Figura 4.11: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en la mezcla EtOH:H₂O (85:15). Cromatogramas correspondientes al blanco de enzima, blanco de sustratos y muestra de reacción, luego de 3 horas de reacción.

Los cromatogramas obtenidos para el blanco de enzima y el tiempo de reacción son coincidentes (Figura 4.11). Se observan dos picos principales: el de mayor tiempo de retención (t_R 14 minutos) que corresponde al sustrato, y el de 5 minutos de tiempo de retención, que se correspondería con el sustrato hidrolizado. No se observa algún pico de tiempo de retención intermedio asociado a la formación de Bz-Arg-Glyol.

En resumen, utilizando este segundo protocolo de reacción, araujaína no podría catalizar la formación de Bz-Arg-Glyol en ninguno de los dos medios ensayados.

Como se menciona en 1.1.10.2., la síntesis de isómeros peptídicos se basa en la capacidad de las peptidasas de acilar compuestos distintos a aminoácidos, así como emplear donores de acilo diferentes a los aminoácidos codificados. Se ha encontrado en bibliografía que diversas cisteín peptidasas vegetales del tipo de la papaína fueron buenos biocatalizadores para condensar derivados de Arginina con alquilaminas. Araujiaína y funaistraína demostraron tener un rendimiento en síntesis de derivados Arg-alquilamida similares o superiores al de papaína (70%) (Morcelle *et al.*, 2009). Papaína y araujiaína fueron ensayadas en distintas configuraciones (libres y adsorbidas sobre soportes) y distintos medios acuoso-orgánicos como catalizadores en la condensación de Z-Ala con glicinol, con resultados exitosos. Utilizando papaína inmovilizada en ACN con 1% y 4% de agua, bajo control termodinámico se reportó un 70% de Z-Ala-Glyol luego de 72 horas de reacción y con araujiaína libre en buffer Na_2CO_3 - NaHCO_3 de pH 10 bajo control cinético, 60% luego de 6 horas de reacción (Morcelle *et al.*, 2008).

Por lo expuesto anteriormente se esperaba que las reacciones efectuadas empleando Bz-Arg-OEt, y glicinol como acil donador, tuvieran resultados positivos. Dentro de las posibles causas que podrían justificar este resultado inesperado, podrían considerarse la cantidad de TEA y de enzima utilizadas. En la síntesis bajo control cinético, es deseable que el nucleófilo se encuentre en su forma básica. Quizás la cantidad de TEA empleada no haya sido suficiente para lograr esta condición. Por otro lado, como parte de la eventual continuación de este trabajo, podría ensayarse esta reacción empleando cantidades superiores de enzima y otro tipo de soporte, en caso de utilizar nuevamente araujiaína inmovilizada.

4.1.4. Síntesis enzimática de derivados de Bz-Arg-Gly utilizando glicina como nucleófilo. Seguimiento por TLC y RP-HPLC

Según lo detallado en la sección 3.1.7., se realizó un nuevo ensayo de síntesis, empleando glicina como nucleófilo y una mezcla ACN:H₂O 85:15 como medio de reacción. Los resultados obtenidos por TLC se muestran en la Figura 4.12.

TLC

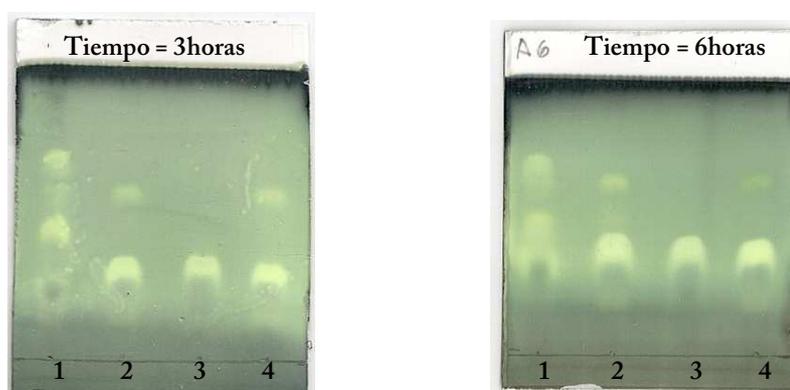


Figura 4.12: Análisis por TLC de la reacción Bz-Arg-OEt con glicina en la mezcla ACN:H₂O (85:15). Puntos de siembra 1: patrón de sustrato (Bz-Arg-OEt) y patrón de sustrato hidrolizado (Bz-Arg-OH). Puntos de siembra 2: blanco de enzima. Puntos de siembra 3: blanco de sustratos. Puntos de siembra 4: muestras de reacción correspondientes.

En los puntos de siembra número 1 fueron sembrados 3 μ l del patrón de sustrato y 3 μ l del patrón de hidrólisis.

Para ambos tiempos, la muestra de reacción reveló solo las manchas correspondientes al sustrato ($R_f=0,58$), al producto de hidrólisis ($R_f=0,45$) y a impurezas ($R_f=0,32$).

Los cromatogramas obtenidos por HPLC para la muestra de reacción y los blancos correspondientes se muestran en la figura 4.13 (en el anexo 8.1.3. se encuentran las gráficas individuales).

RP-HPLC

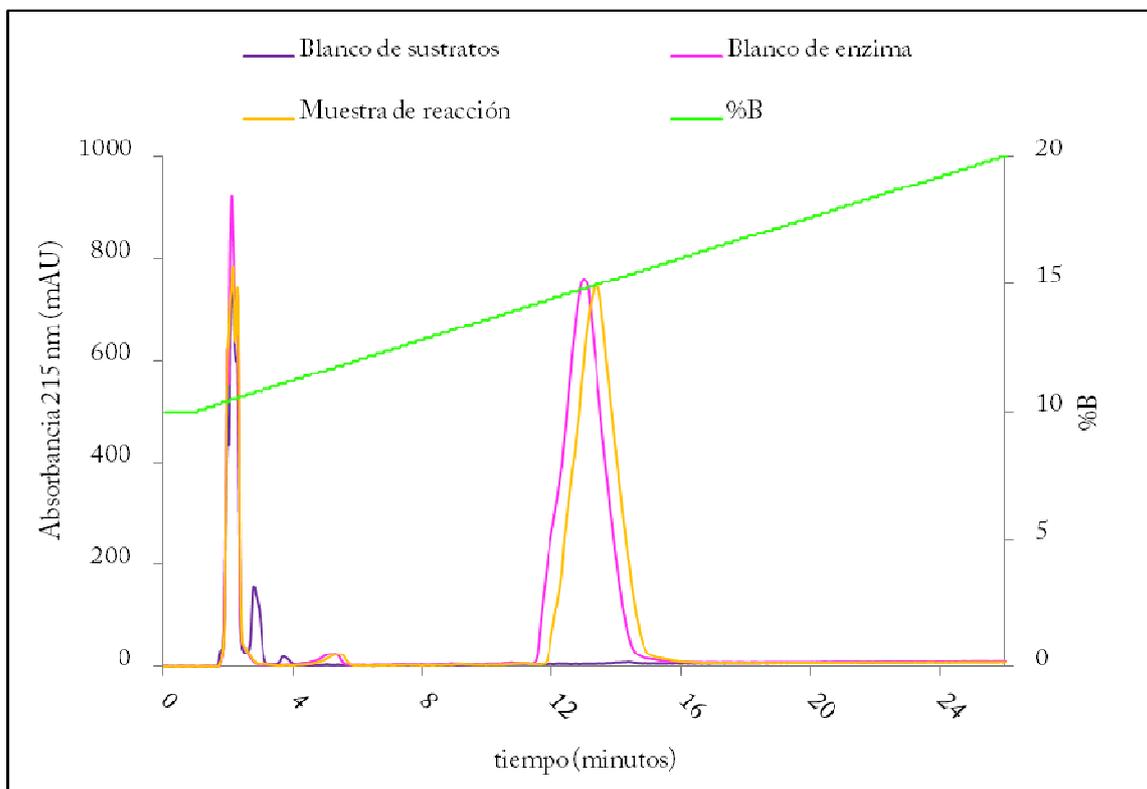


Figura 4.13: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicina en la mezcla ACN:H₂O (85:15). Cromatogramas correspondientes al blanco de enzima, blanco de sustratos y muestra de reacción, luego de 3 horas de reacción.

Los perfiles obtenidos para el blanco de enzima y la muestra de reacción son prácticamente idénticos. Se observa el pico correspondiente al sustrato (t_R 14 minutos) y el de su producto de hidrólisis (t_R 5,10 minutos). Nuevamente, no se observó la presencia de un pico de tiempo de retención intermedio que pudiera asociarse a la formación del producto esperado (Bz-Arg-Gly) en las condiciones ensayadas.

En el año 2005, Hou *et al* lograron la unión de Bz-Arg-OEt a Gly-Asp-(NH₂)₂ mediante el uso de alcalasa y bajo control cinético en disolventes orgánicos. Ellos optimizaron las condiciones de reacción estudiando los efectos de varios factores, incluyendo el contenido de agua, temperatura, pH y el tiempo de reacción. En ese trabajo, los investigadores emplearon mezclas orgánico:acuosas 85:15, y ensayaron como disolventes orgánicos etanol, metanol, dimetilformamida (DMF), dimetilsulfóxido (DMSO), acetato de etilo, cloroformo y acetonitrilo. Los rendimientos obtenidos fueron superiores al 50% en EtOH, DMF y MeOH, siendo el medio que contenía EtOH el que mostro mayor rendimiento (73,6% después de 8 horas de reacción). Por otro lado, aunque este grupo utilizó una proteasa serinica, se esperaba poder reproducir alguno de sus resultados, dado que este tipo de enzimas forman un intermediario acil-enzima covalente al igual que las empleadas en este trabajo.

4.1.5. Síntesis enzimática de derivados de Bz-Arg-Ala-OMe utilizando Ala-OMe como nucleófilo. Seguimiento por TLC y RP-HPLC

En este nuevo ensayo de síntesis (véase 3.1.8.), se empleo araujaína libre, y Ala-OMe como nucleófilo. Se ensayaron como medios de reacción mezclas organico:acuosas 85:15. Los resultados obtenidos para la mezcla conteniendo ACN se discuten en primer lugar (Figuras 4.14-4.16), y en segundo lugar se analizan los obtenidos para la mezcla que contiene EtOH (Figura 4.17) como solvente organico. En el anexo 8.1.4. se encuentran los resultados obtenidos mediante RP-HPLC en graficas individuales.

Medio ACN:H₂O (85:15)

TLC

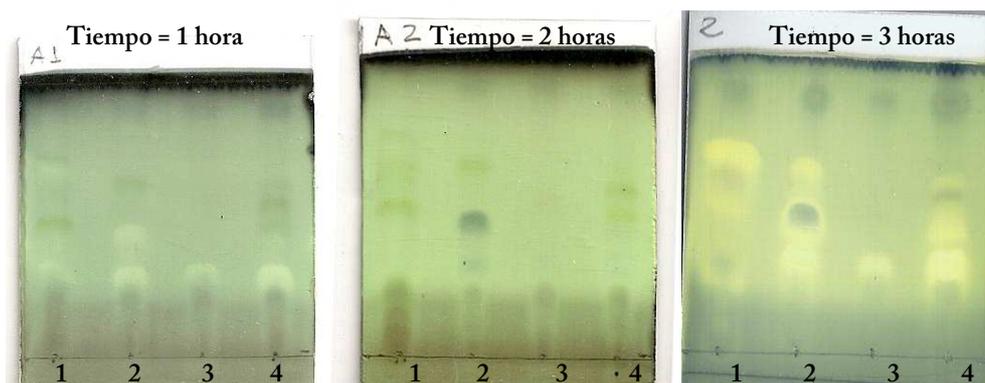


Figura 4.14: Análisis por TLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla ACN:H₂O (85:15). Puntos de siembra 1: patron de sustrato (Bz-Arg-OEt) y patron de sustrato hidrolizado (Bz-Arg-OH). Puntos de siembra 2: blanco de enzima. Puntos de siembra 3: blanco de sustratos. Puntos de siembra 4: muestras de reacción.

Las corridas correspondientes al blanco de enzima y a la muestra de reacción revelan las manchas correspondientes al sustrato ($R_f=0,56$) y su producto de hidrólisis ($R_f=0,45$). Sin embargo, en la muestra de reacción se evidenciaron una serie de manchas de R_f intermedio (0,5 aproximadamente), incluso en la placa correspondiente a 1 hora de reacción.

Los cromatogramas obtenidos por HPLC para la muestra de reacción y los blancos correspondientes se muestran en la figura 4.15.

RP-HPLC

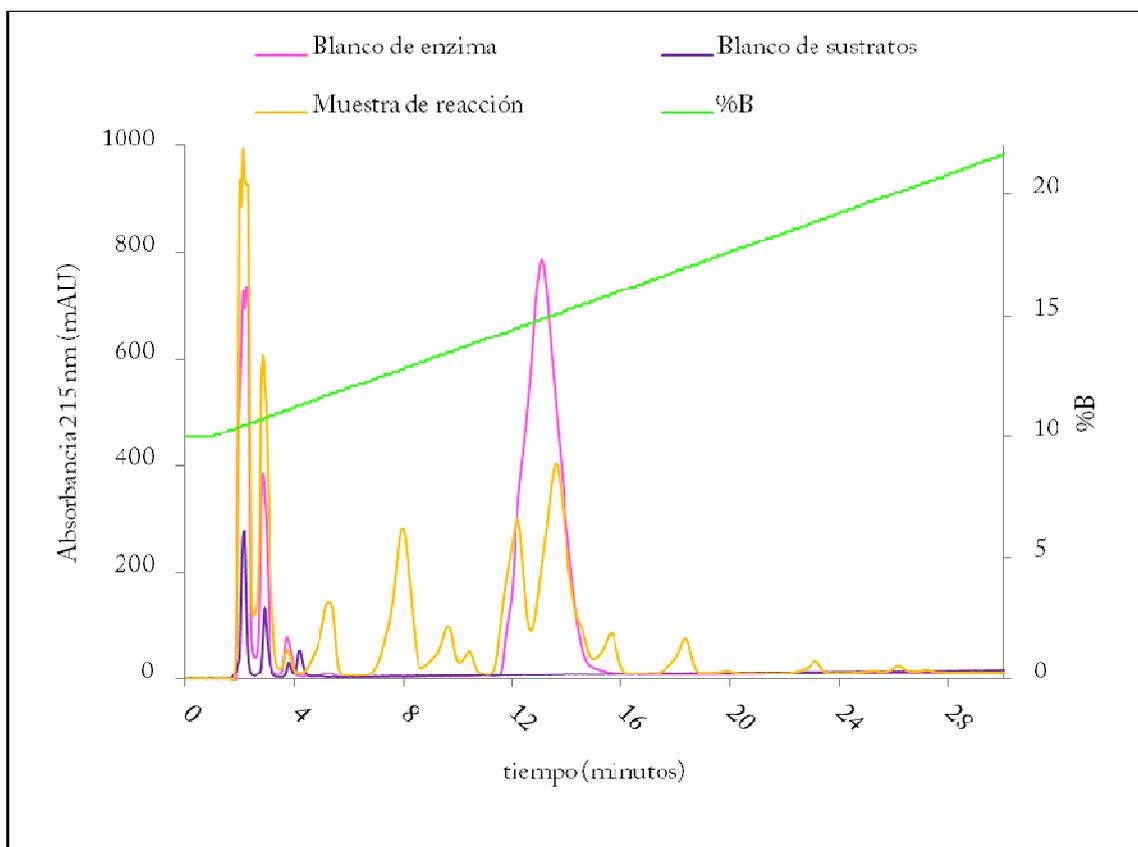


Figura 4.15: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla ACN:H₂O (85:15). Cromatogramas correspondientes al blanco de enzima, blanco de sustratos y muestra de reacción, luego de 3 horas de reacción.

La muestra de reacción tomada luego de 3 horas de reacción presentó varios picos, además de los correspondientes al sustrato (14 minutos) y al producto de hidrólisis (5 minutos). A partir de estos resultados se dedujo que los picos de tiempo de retención intermedio debían estar asociados a la formación del producto esperado (Bz-Arg-Ala-OMe) y de su ácido (Bz-Arg-Ala-OH), en tanto que aquellos picos de t_R mayor a 14 minutos, solo podrían ser resultado de productos de polimerización del tipo Bz-Arg-(Ala)_n-OMe, de naturaleza más hidrofóbica que el sustrato.

En la Figura 4.16 se muestra los cromatogramas obtenidos para muestras de reacción tomadas luego de 1, 3 y 6 horas. Se observa que no existe un aumento relativo significativo de la cantidad de ninguno de los productos, ni disminución relativa de la cantidad sustrato. Esto podría explicarse en base a que antes de la primera hora se habría alcanzado el equilibrio de la reacción.

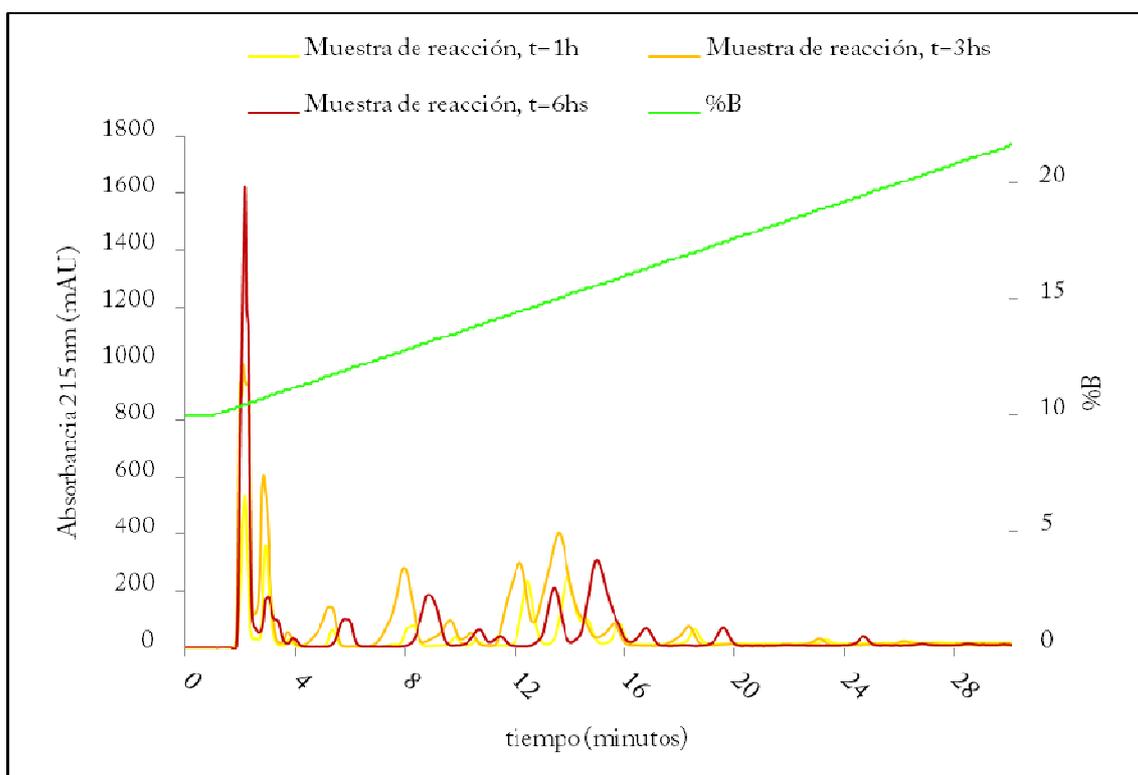


Figura 4.16: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla ACN:H₂O (85:15). Cromatogramas correspondientes a muestras de reacción tomadas luego de 1, 2 y 3 horas de reacción.

Medio EtOH:H₂O (85:15)

Para la reacción ensayada en la mezcla EtOH:H₂O se realizó únicamente el análisis por RP-HPLC. Los resultados se muestran en la Figura 4.17, y los cromatogramas individuales se encuentran en el anexo 8.1.4.

RP-HPLC

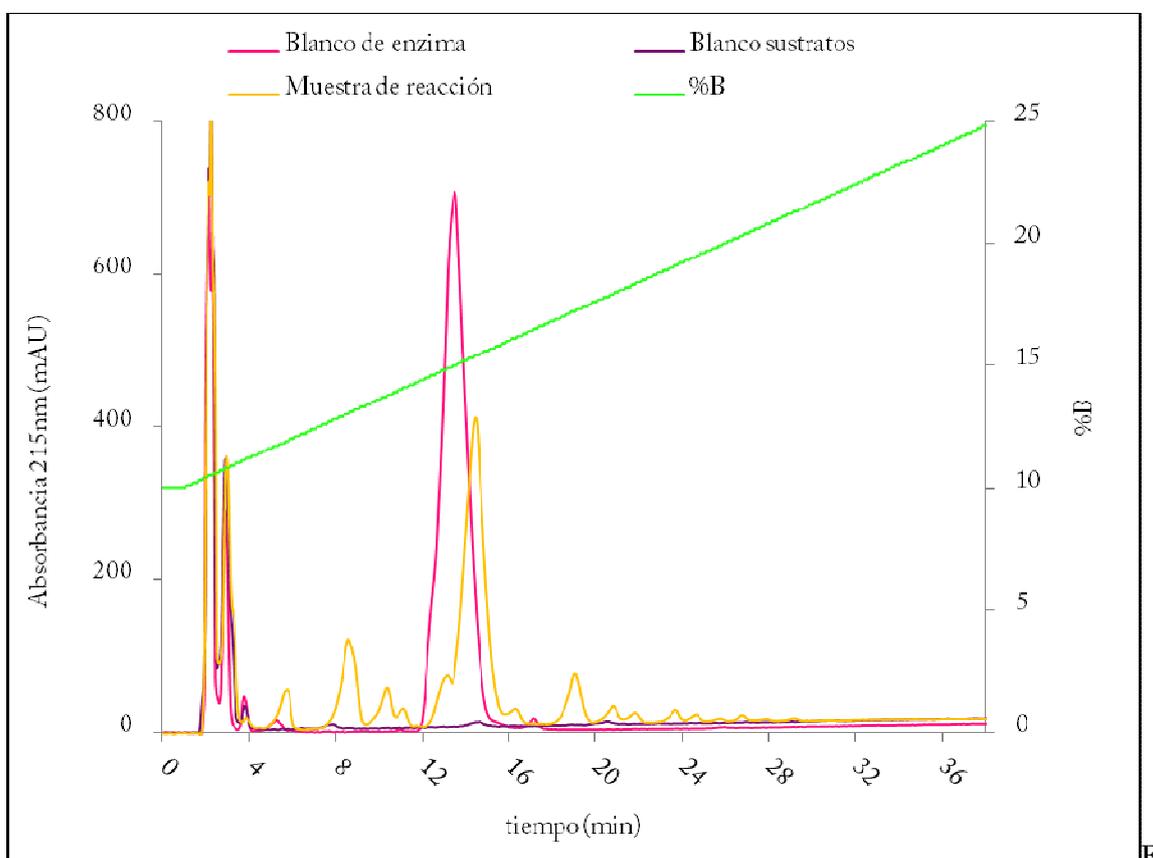


figura 4.17: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla EtOH:H₂O (85:15). Cromatogramas correspondientes al blanco de enzima, blanco de sustratos y muestra de reacción, luego de 3 horas de reacción.

El perfil obtenido para la muestra de reacción es muy similar al observado en la figura 4.15 (correspondiente a la misma reacción, ensayada en la mezcla ACN:H₂O). Nuevamente

podemos observar los picos correspondientes al sustrato (14 minutos) y su producto de hidrólisis (5 minutos), y la misma serie de picos de tiempos de retención intermedio. También se observaron picos de tiempo de retención mayor al del sustrato.

En la figura 4.18 se muestran los cromatogramas correspondientes a las muestras de reacción tomadas luego de 1, 3 y 6 horas.

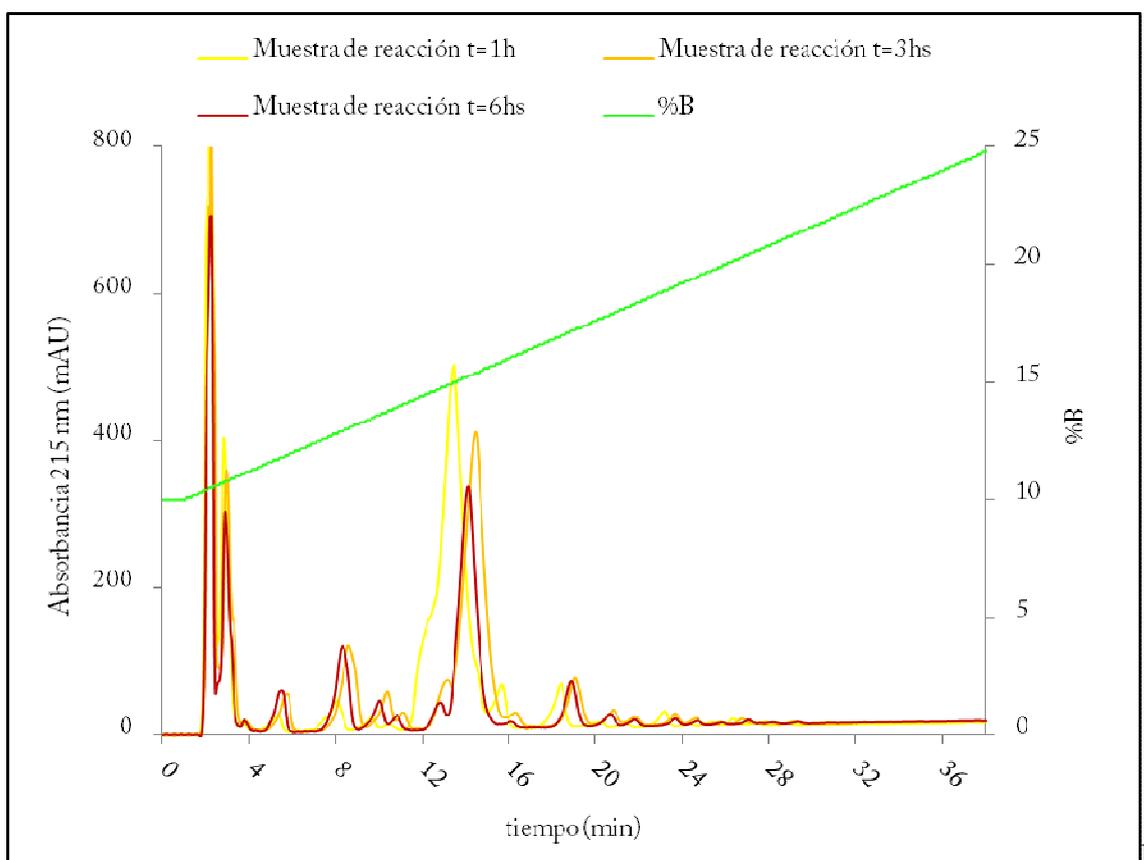


figura 4.18: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla EtOH:H₂O (85:15). Cromatogramas correspondientes a muestras de reacción tomadas luego de 1, 3 y 6 horas de reacción.

Los cromatogramas obtenidos para los tres tiempos de reacción son muy similares, aunque puede apreciarse (principalmente entre las muestras tomadas luego de 1 y 3 horas de reacción) una disminución gradual en la cantidad de sustrato, acompañado de un pequeño aumento de los picos de tiempo de retención intermedios (8, 9,5 y 10,5 minutos).

Por otro lado, la figura 4.19 muestra la superposición de los cromatogramas obtenidos para las muestras de reacción, de ambos medios, tomadas luego de 3 horas de incubación. Puede observarse que el perfil de picos es similar en ambos medios, pero en EtOH los picos se retrasan. Por otro lado, la altura y el área de los picos que se corresponderían con los productos mas simples (Bz-Arg-Ala-OMe y Bz-Arg-(Ala)₂-OMe), parecen ser mayores en la muestra tomada del medio de reacción que contiene ACN. Sin embargo también puede apreciarse que en el cromatograma obtenido para la reacción efectuada en EtOH, aparecen mas picos, pequeños y de mayores tiempos de retención. Esto podría estar indicando que los rendimientos para los diferentes productos de condensación, son diferentes en los diferentes medios. Según lo observado, los rendimientos para los productos de condensación mas simples serían mayores en ACN, mientras que los rendimientos para los productos poliméricos (Bz-Arg-(Ala)_n-OMe) serían mayores en EtOH. Otra explicación posible es que la reacción ocurre mas rapidamente en EtOH, por lo que parte de la Bz-Arg-Ala-OMe y la Bz-Arg-(Ala)₂-OMe que se producirían a tiempos cortos de reacción, estarían siendo nuevamente utilizados como sustrato para reacciones de condensación sucesivas, lo que daría lugar a mayor cantidad de productos de tipo Bz-Arg-(Ala)_n-OMe

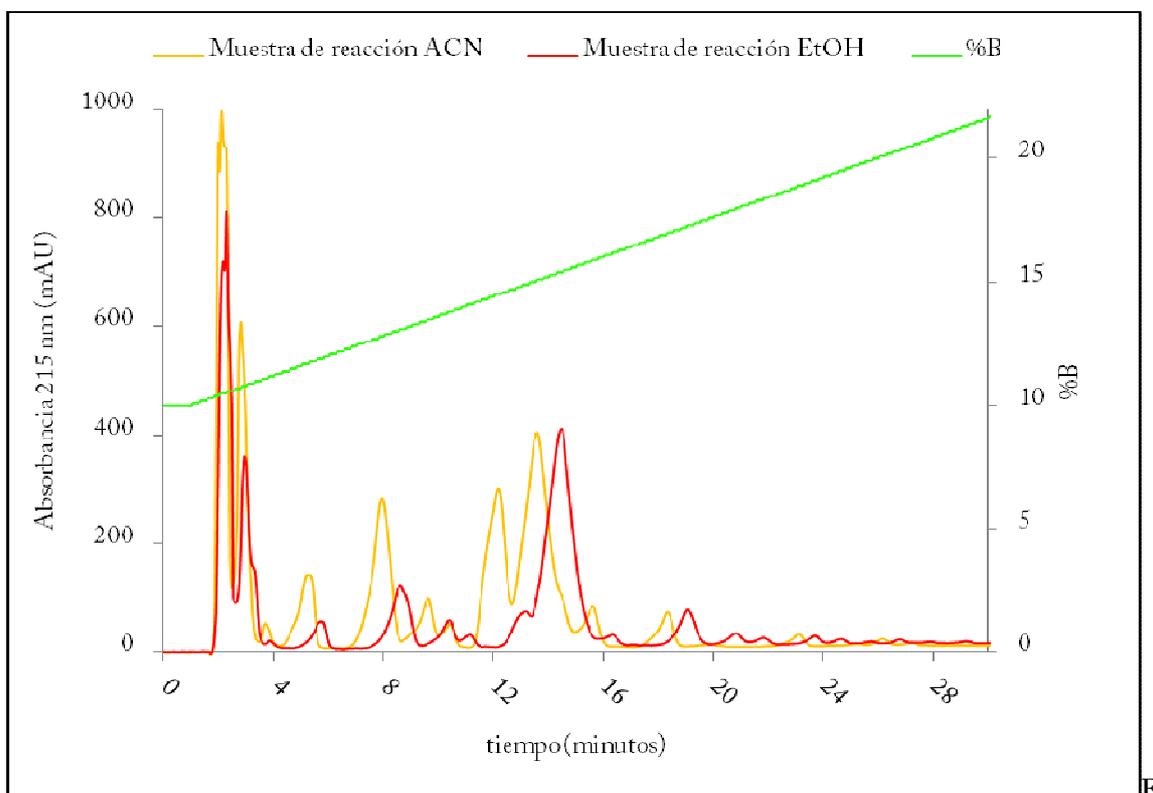


figura 4.19: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe. Cromatogramas correspondientes a muestras de reacción tomadas luego 3 horas de reacción, en las mezclas ACN:H₂O y EtOH:H₂O

4.1.6. Análisis de los productos obtenidos mediante HPLC-ESI MS

Con el objetivo de conocer en forma precisa la naturaleza de los picos obtenidos en la reacción efectuada empleando Ala-OMe como nucleófilo, una muestra de esta reacción efectuada en la mezcla ACN:H₂O, tomada luego de 3 horas, fue analizada por HPLC-MS electrospray. Las condiciones de análisis utilizadas fueron detalladas en la sección 3.1.9. En la Figura 4.20 se muestra el cromatograma obtenido (el reporte completo se encuentra en el anexo 8.1.5.)

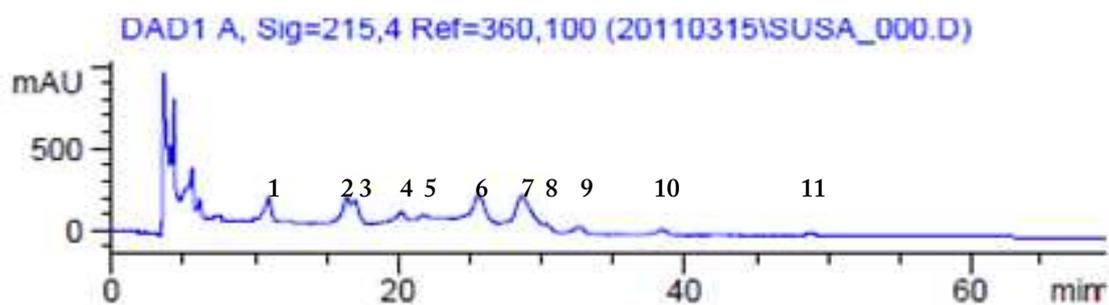


Figura 4.20: Análisis por HPLC-MS. Cromatograma correspondiente a la muestra de reacción tomada luego 3 horas de reacción, en la mezcla ACN:H₂O.

En el cromatograma obtenido, todos los tiempos de retención corresponden al doble de los tiempos obtenidos de la Figura 4.15. Esto es razonable ya que en este análisis la velocidad de fase móvil empleada fue la mitad de la utilizada en las corridas realizadas con anterioridad. Claramente puede observarse un pico de 10,9 minutos de tiempo de retención (pico número 1) que, de acuerdo a los resultados hasta aquí analizados, debiera corresponder a la Bz-Arg-OH. También podemos distinguir el que se correspondería con el sustrato (Bz-Arg-OEt) y cuyo tiempo de retención es de 28,6 minutos (pico número 8).

El análisis por espectrometría de masas electrospray de cada uno de los picos obtenidos en el cromatograma (Figura 4.20), en orden creciente de tiempos de retención, se muestra en las figuras 4.21-4.31.

Por otro lado, en la tabla 4.5 se resumen las masas exactas y las relaciones m/z para el sustrato, su producto de hidrólisis, y los posibles productos de síntesis, calculadas empleando el software ChemDraw Pro 8.0®.

Tabla 4.5: Masas exactas y las relaciones m/e esperadas para los distintos compuestos esperados en la muestra de reacción.

Compuesto	Masa exacta	m/z
Bz-Arg-OEt	306,17	306,17 (100,0%)
		307,17 (18,3%)
		308,18 (1,4%)
Bz-Arg-OH	278,14	278,14 (100,0%)
		279,14 (14,8%)
		280,14 (1,8%)
		279,13 (1,5%)
Bz-Arg-Ala-OMe	363,19	363,19 (100,0%)
		364,19 (20,9%)
		365,20 (1,8%)
		365,19 (1,2%)
Bz-Arg-Ala-OH	349,18	349,18 (100,0%)
		350,18 (18,3%)
		351,18 (2,7%)
		350,17 (1,9%)
Bz-Arg-(Ala) ₂ -OMe	434,23	434,23 (100,0%)
		435,23 (22,9%)
		436,23 (3,9%)
		435,22 (2,2%)
Bz-Arg-(Ala) ₂ -OH	420,46	420,21 (100,0%)
		421,22 (21,8%)
		422,22 (3,3%)
		421,21 (2,2%)

Tabla 4.5 (continuación): Masas exactas y las relaciones m/e esperadas para los distintos compuestos esperados en la muestra de reacción.

Bz-Arg-(Ala) ₃ -OMe	505,26	505,26 (100,0%) 506,27 (25,5%) 507,27 (5,0%) 506,26 (2,6%)
Bz-Arg-(Ala) ₃ -OH	491,25	491,25 (100,0%) 492,25 (26,6%) 493,26 (2,8%) 493,25 (1,9%)
Bz-Arg-(Ala) ₄ -OMe	576,3	576,30 (100,0%) 577,31 (28,8%) 578,31 (5,5%) 577,30 (3,0%)
Bz-Arg-(Ala) ₄ -OH	562,29	562,29 (100,0%) 563,29 (27,7%) 564,29 (5,8%) 563,28 (3,0%)
Bz-Arg-(Ala) ₅ -OMe	647,34	647,34 (100,0%) 648,34 (35,0%) 649,35 (5,0%) 649,34 (2,7%) 650,35 (1,0%)
Bz-Arg-(Ala) ₅ -OH	633,32	633,32 (100,0%) 634,33 (31,1%) 635,33 (6,3%) 634,32 (3,3%) 636,33 (1,1%)

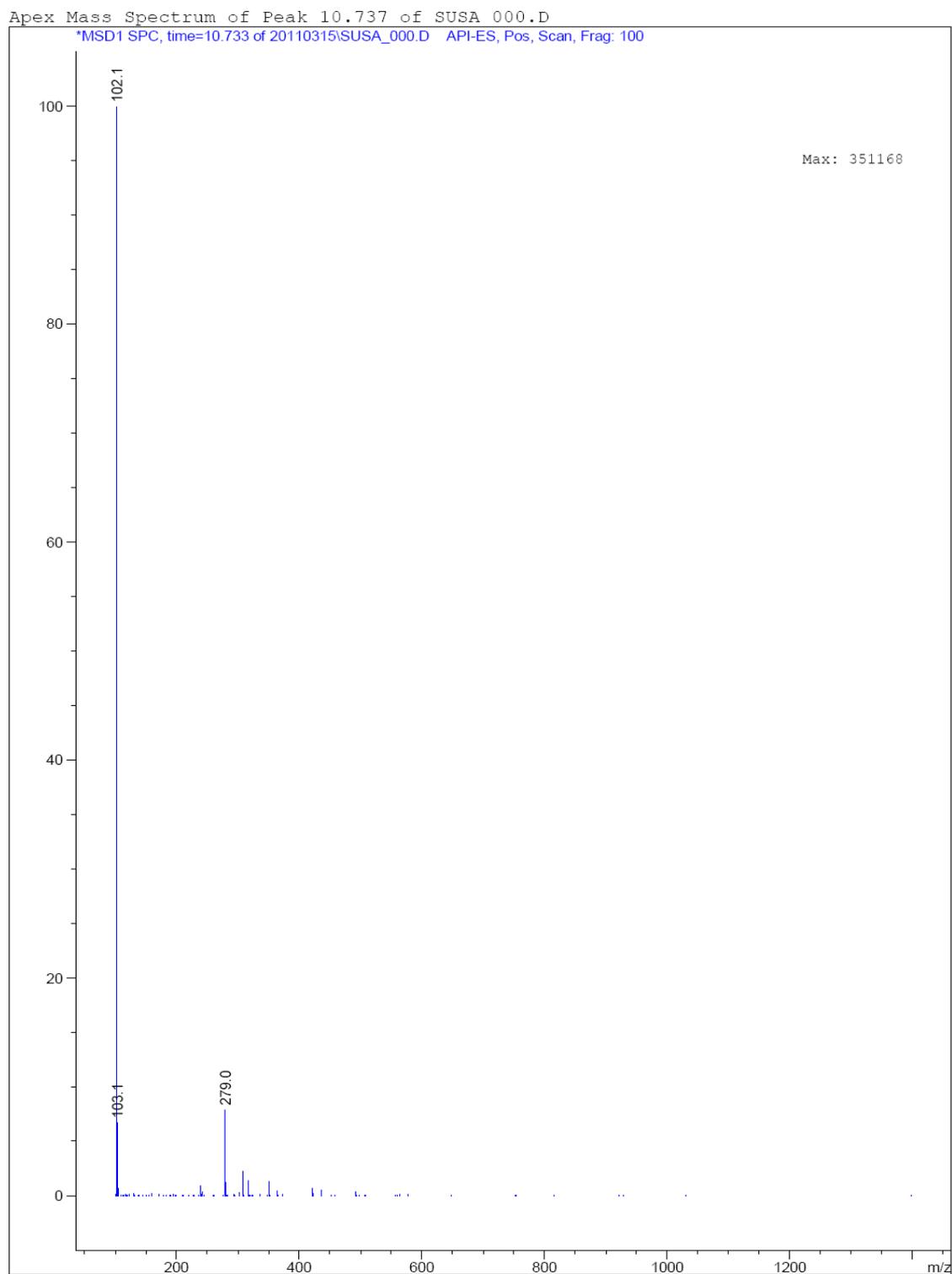


Figura 4.21: Espectro de masas obtenido para el pico número 1 (t_R : 10,9 minutos)

Los resultados obtenidos se condicinen con lo esperado. A t_R 10,9 minutos aparece una señal correspondiente a una relación m/z igual a 279,0, lo que revela la presencia de Bz-Arg-OH ($M+H^+$).

La señal de relación m/z igual a 102,1 y 103,1 no se corresponde con el doble protonado y no ha sido posible atribuirle un producto que se corresponda con dichos valores.

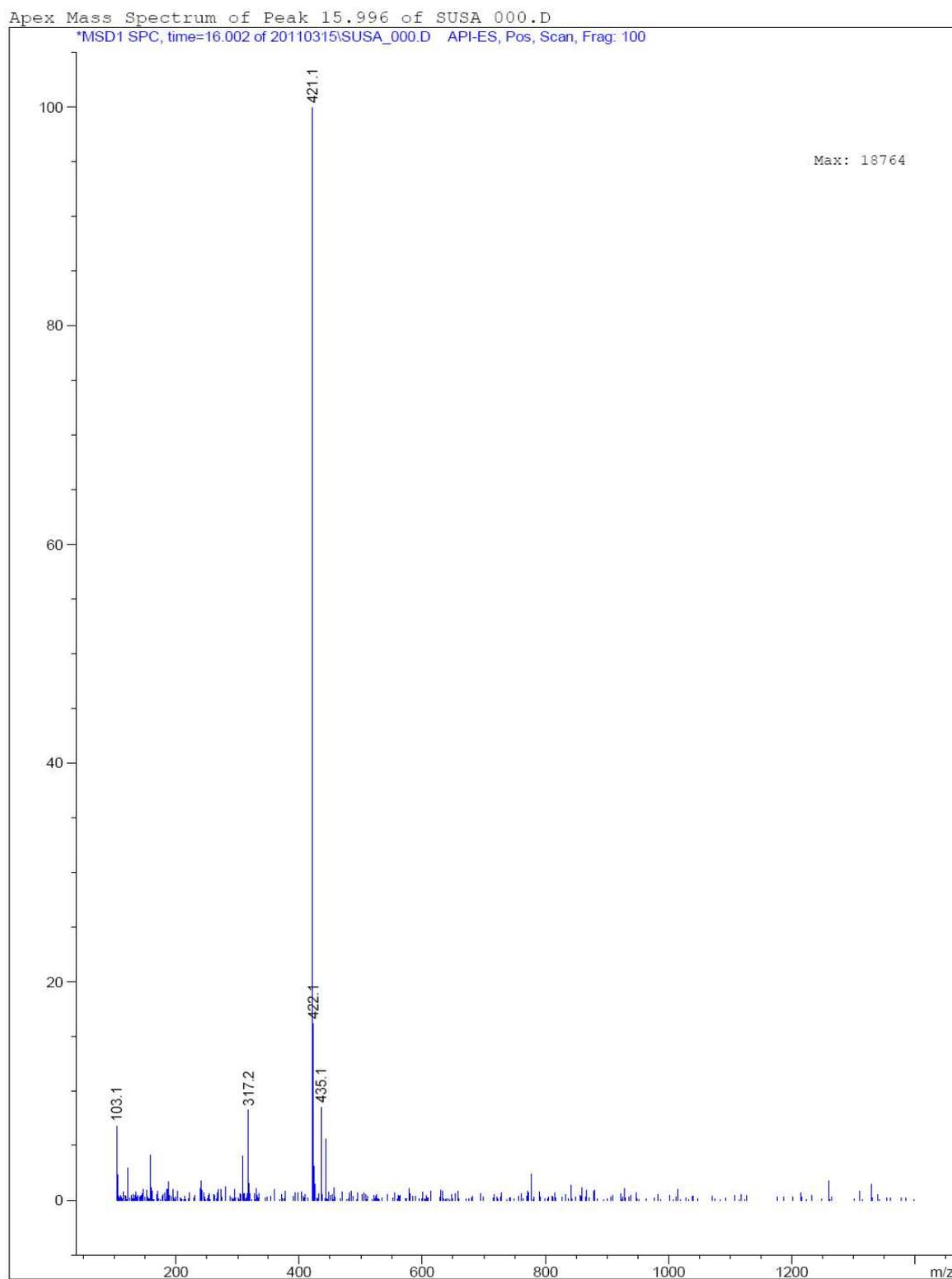


Figura 4.22: Espectro de masas obtenido para el pico número 2 (t_R : 16,3 minutos)

El espectro de masas correspondiente al pico 2 (t_R 16,3 minutos) muestra una señal principal correspondiente a una relación m/z igual a 421,1 y 422,1. Ambos se corresponden con la protonación simple de la Bz-Arg-(Ala)₂-OH ($M+H^+$). También pueden observarse otras dos señales (m/z igual a 103,1 y 317,2) que no se corresponden con ningún otro estado posible de protonación de este compuesto y a las cuales no ha sido posible atribuirle una estructura química.

Por otro lado, también está presente la señal correspondiente a la especie Bz-Arg-(Ala)₂-OMe ($M+H^+$) (m/z igual a 435,1).

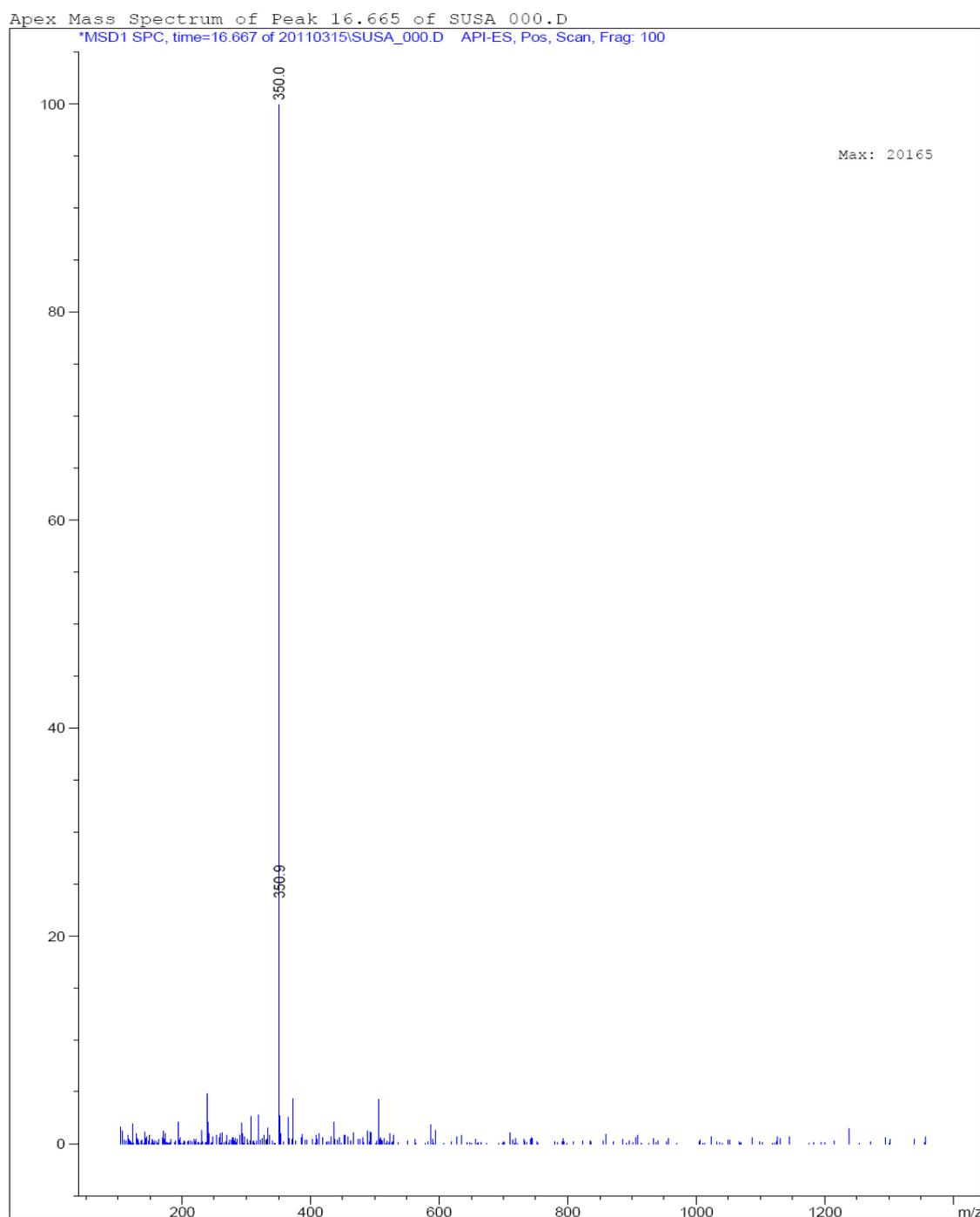


Figura 4.23: Espectro de masas obtenido para el pico número 3 (t_R : 16,9 minutos)

En este espectro, las señales obtenidas se corresponden con un único compuesto: Bz-Arg-Ala-OH ($M+H^+$) (m/z igual a 350,0 y 350,9)

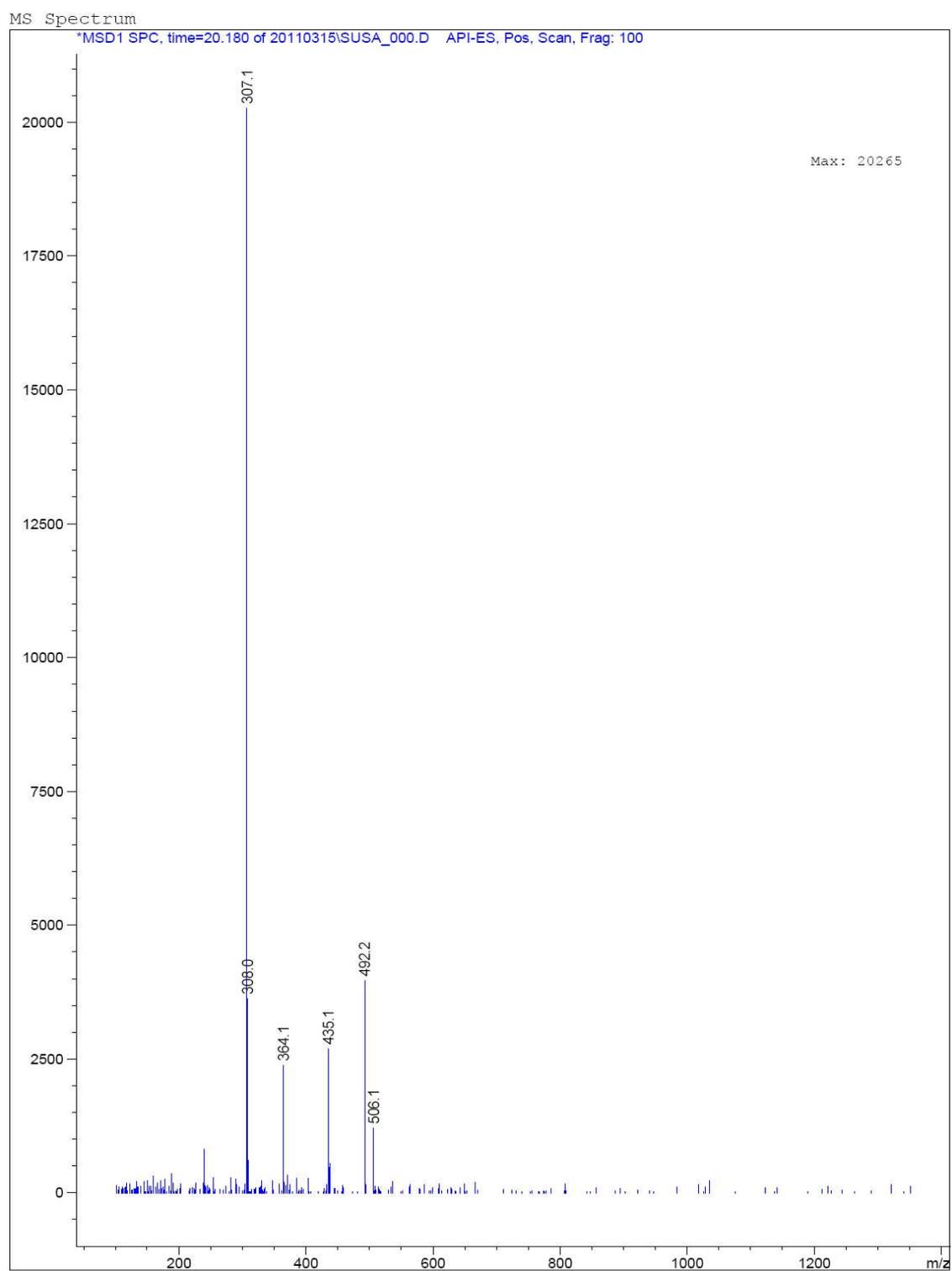


Figura 4.24: Espectro de masas obtenido para el pico número 4 (t_R : 20,18 minutos)

La Figura 4.24 muestra el espectro obtenido para el pico cuyo tiempo de retención es 20,18 minutos. Claramente el mismo es resultado de una mezcla de varios productos. La señal principal revela la presencia de sustrato (Bz-Arg-OEt, M+H⁺ igual a 307,1 y 308,0). Por otro lado, las señales de menor intensidad corresponden a Bz-Arg-Ala-OMe (M+H⁺) (m/z igual a 364,1), Bz-Arg-(Ala)₂-OMe (M+H⁺) (m/z igual a 435,1), Bz-Arg-(Ala)₃-OH (M+H⁺) (m/z igual a 492,2) y Bz-Arg-(Ala)₃-OMe (M+H⁺) (m/z igual a 506,1)

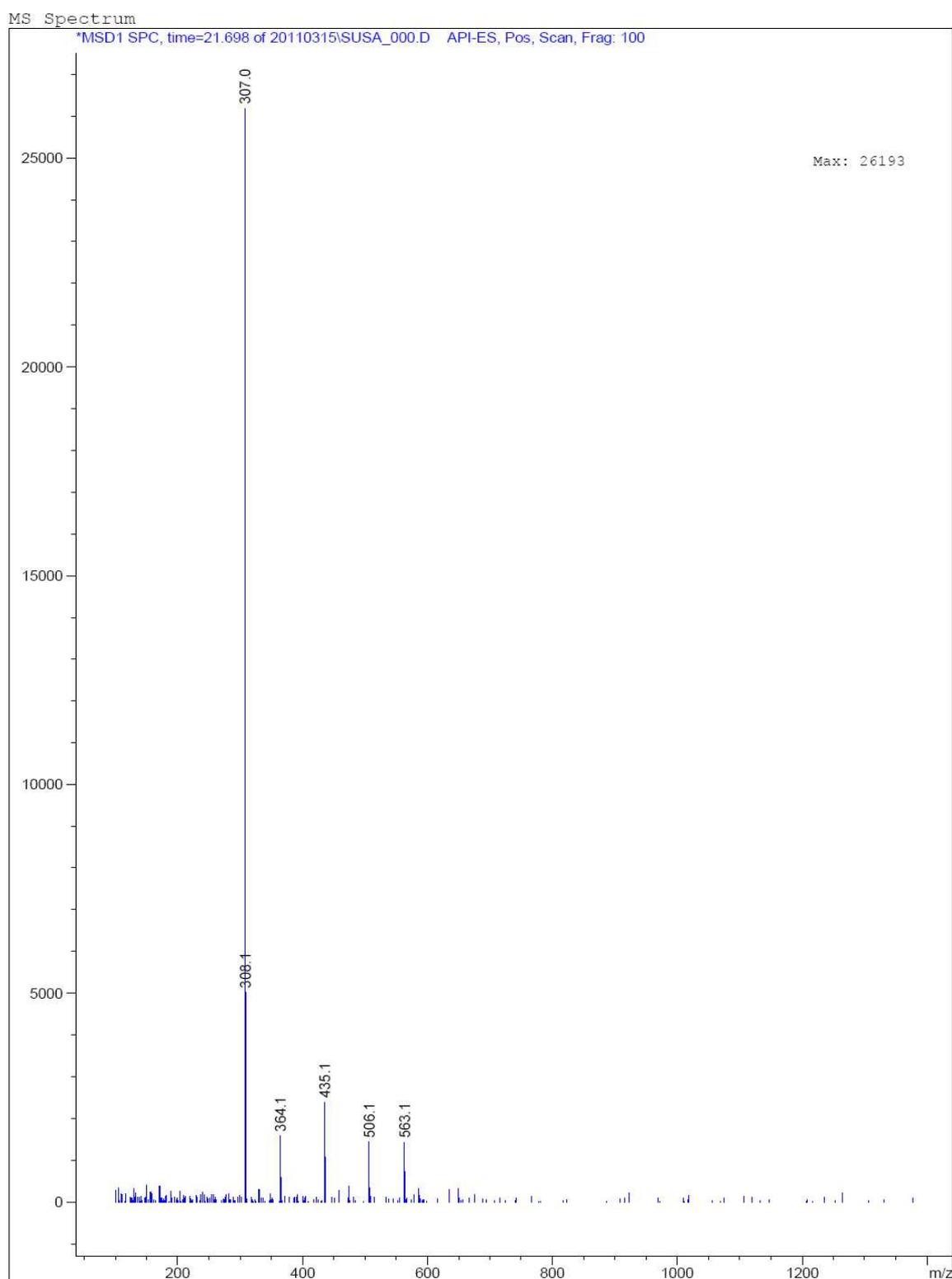


Figura 4.25: Espectro de masas obtenido para el pico número 5 (t_R : 21,7 minutos)

En este espectro (Figura 4.25), nuevamente puede observarse una mezcla de compuestos. La señal principal corresponde al sustrato (Bz-Arg-OEt, $M+H^+$ igual a 307,0 y 308,1), y las señales secundarias de menor intensidad corresponden a Bz-Arg-Ala-OMe ($M+H^+$) (m/z igual a 364,1), Bz-Arg-(Ala)₂-OMe ($M+H^+$) (m/z igual a 435,1) y Bz-Arg-(Ala)₃-OMe ($M+H^+$) (m/z igual a 506,1) y Bz-Arg-(Ala)₄-OH ($M+H^+$) (m/z igual a 563,1)

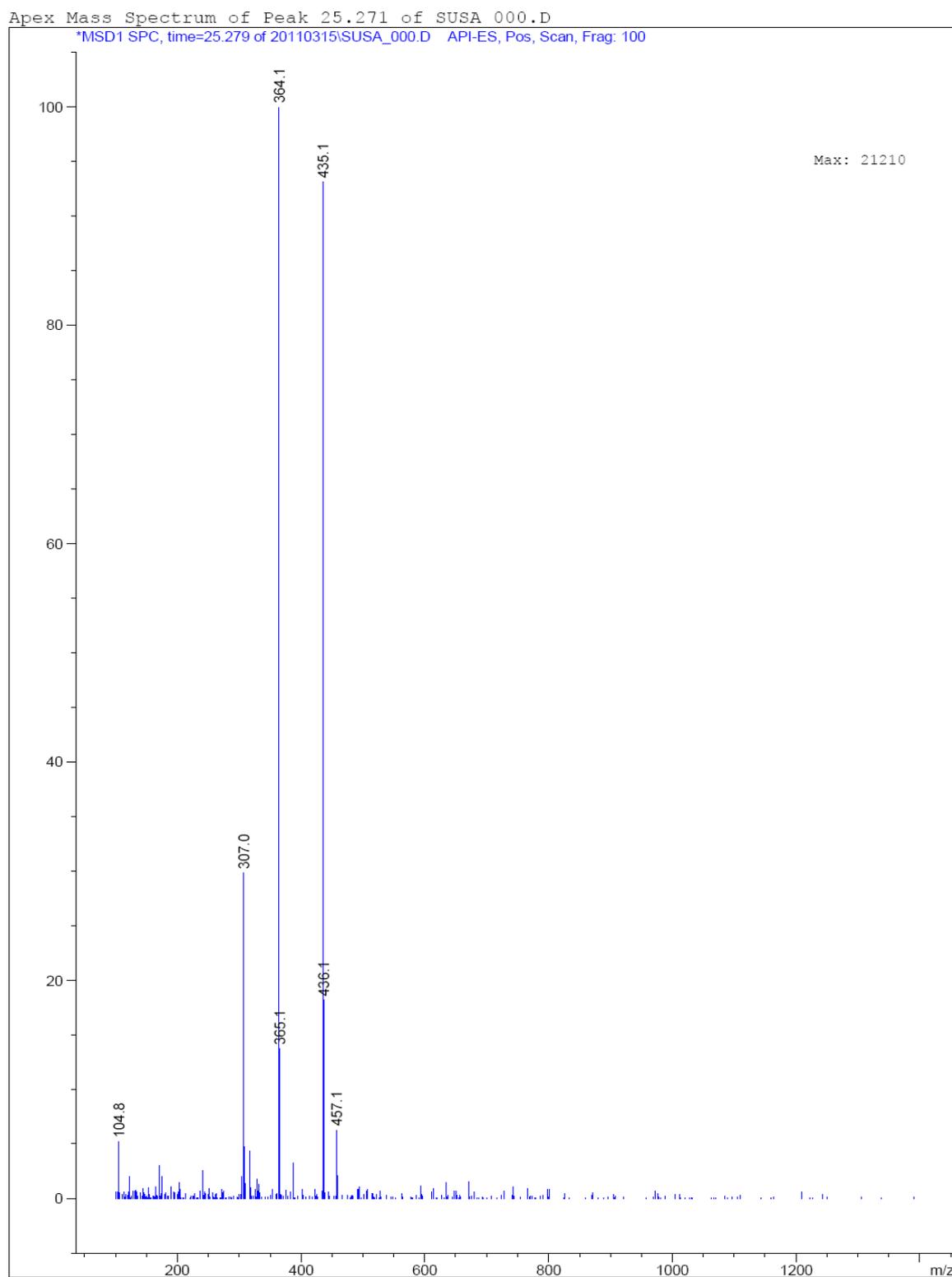


Figura 4.26: Espectro de masas obtenido para el pico número 6 (tr: 25,3 minutos)

En la figura 4.26., las señales obtenidas corresponden a diferentes compuestos. La señal principal, cuya relación m/z es 364,1 y 365,1, se corresponde con nuestro producto de síntesis esperado: la Bz-Arg-Ala-OMe ($M+H^+$). También pueden observarse las señales correspondientes al producto de dos condensaciones sucesivas del sustrato con el nucleófilo (Bz-Arg-(Ala)₂-OMe; $M+H^+$ igual a 435,1 y 436,1) y al sustrato (Bz-Arg-OEt; $M+H^+$ igual a 307,0)

Nuevamente aquí se observan dos señales (m/z igual a 104,8 y 457,1) que no se corresponden con ningún otro estado posible de protonación de los compuestos mencionados y a las cuales no ha sido posible atribuirle una estructura química.

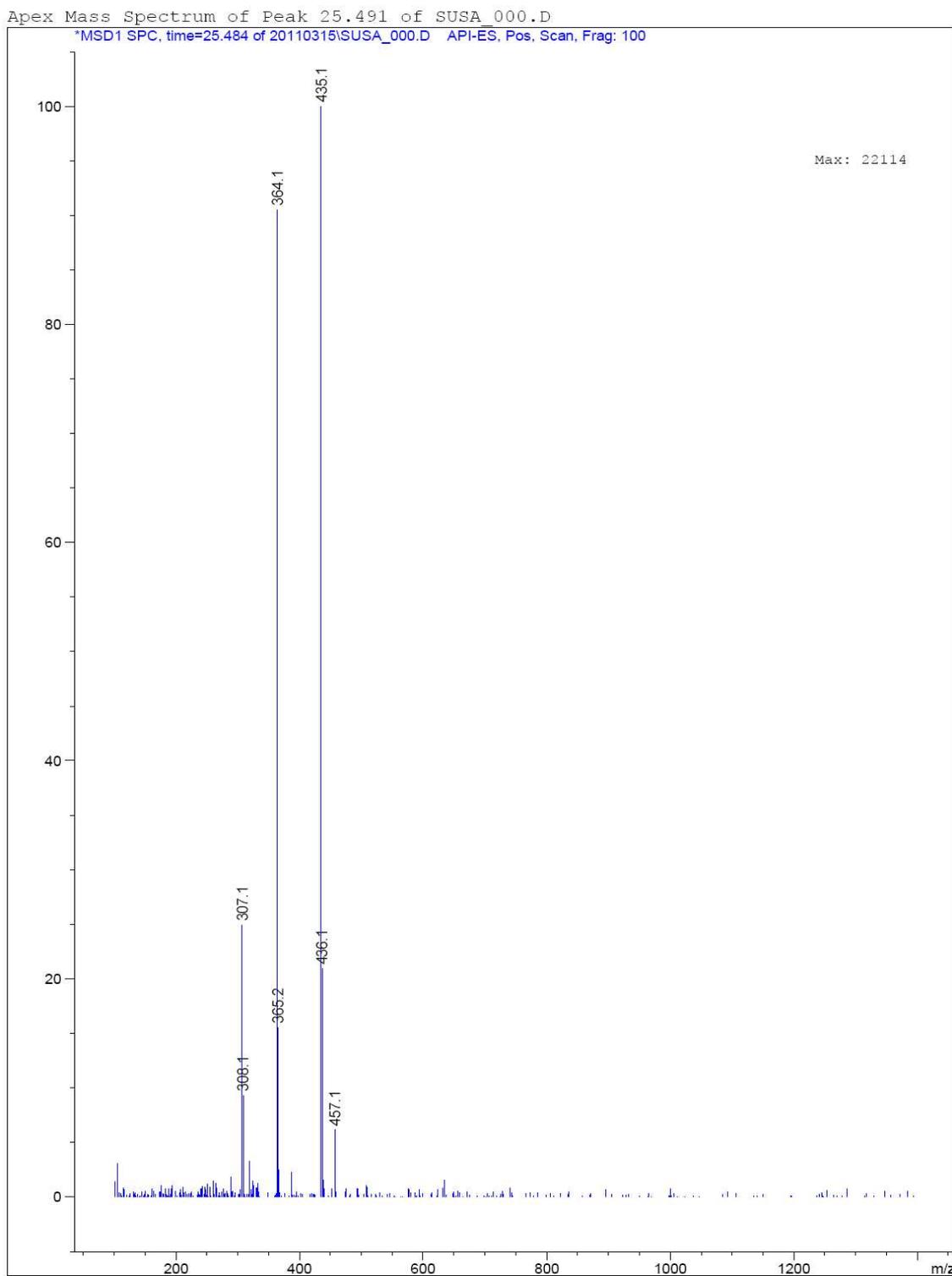


Figura 4.27: Espectro de masas obtenido para el pico número 7 (t_R : 25,5 minutos)

El espectro de masas correspondiente al pico 7 presentó una señal principal que se corresponde con Bz-Arg-(Ala)₂-OMe (M+H⁺ igual a 435,1 y 436,1) y la señal secundaria con Bz-Arg-Ala-OMe (M+H⁺ igual a 364,1 y 365,2). Este resultado y el obtenido en la Figura 4.26 se correlacionan con lo esperado respecto a las polaridades de los posibles productos de reacción. El producto de condensación doble (Bz-Arg-(Ala)₂-OMe) es más hidrofóbico que el producto de reacción más simple (Bz-Arg-Ala-OMe), por lo que este último debe tener un máximo de intensidad a un tiempo menor. Sin embargo es evidente que estas condiciones cromatográficas no permiten la correcta separación de estos compuestos por lo que se encuentra señal intensa de ambos. También pueden observarse las señales correspondientes al sustrato (Bz-Arg-OEt; M+H⁺ igual a 307,0 y 308,1).

Nuevamente aquí se observa la señal correspondiente a m/z igual a 457,1, que no se corresponde con ningún otro estado posible de protonación de los compuestos mencionados y al cual no ha sido posible atribuirle una estructura química.

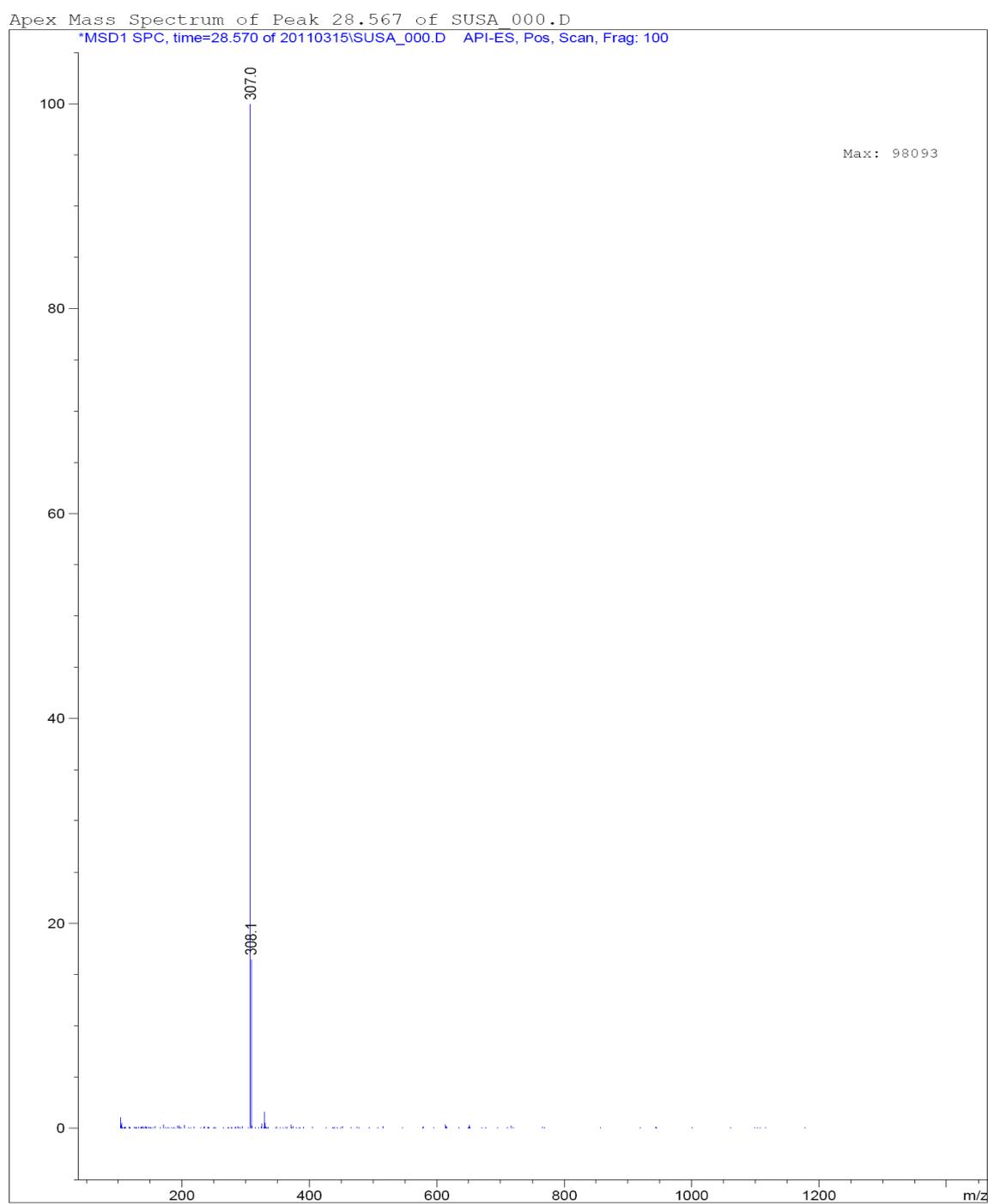


Figura 4.28: Espectro de masas obtenido para el pico número 8 (t_R : 28,6 minutos)

En este espectro, los resultados se condicen con lo esperado ya que las señales obtenidas se corresponden unicamente con el sustrato (Bz-Arg-Ala-OEt, $M+H^+$ igual a 307,0 y 308,1)

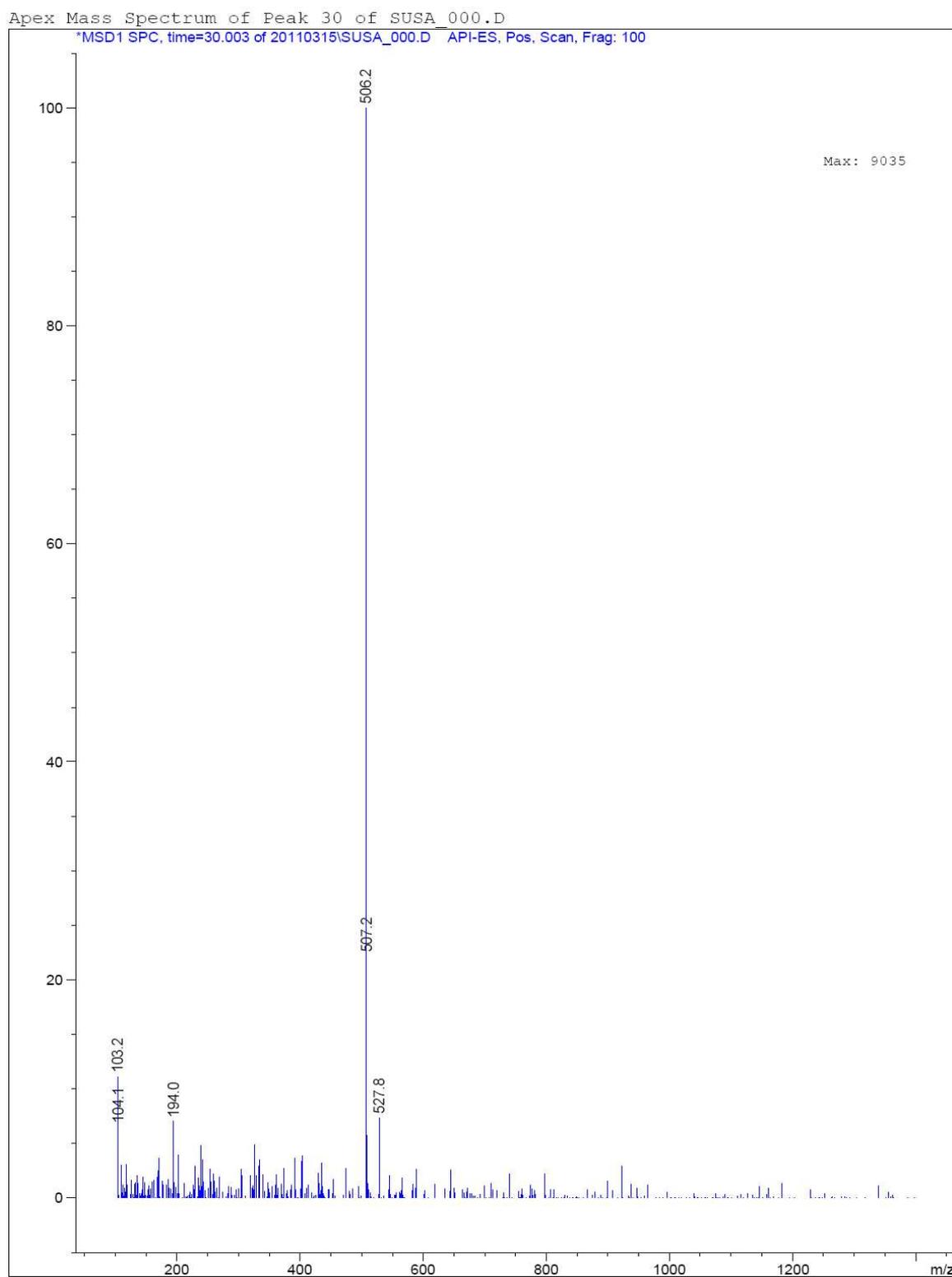


Figura 4.29: Espectro de masas obtenido para el pico número 9 (t_R : 32,6 minutos)

Como esperabamos, se obtuvieron productos de polimerización del tipo Bz-Arg-(Ala)_n-OMe, de naturaleza más hidrofóbica. En este espectro se revela la presencia de Bz-Arg-(Ala)₃-OMe (M+H⁺ igual a 506,2 y 507,2).

También se observan cuatro señales (m/z igual a 103,2, 104,1, 194,0 y 527,8) que no se corresponden con ningún otro estado posible de protonación de la Bz-Arg-(Ala)₃-OMe ni con ningún otro producto de los considerados.

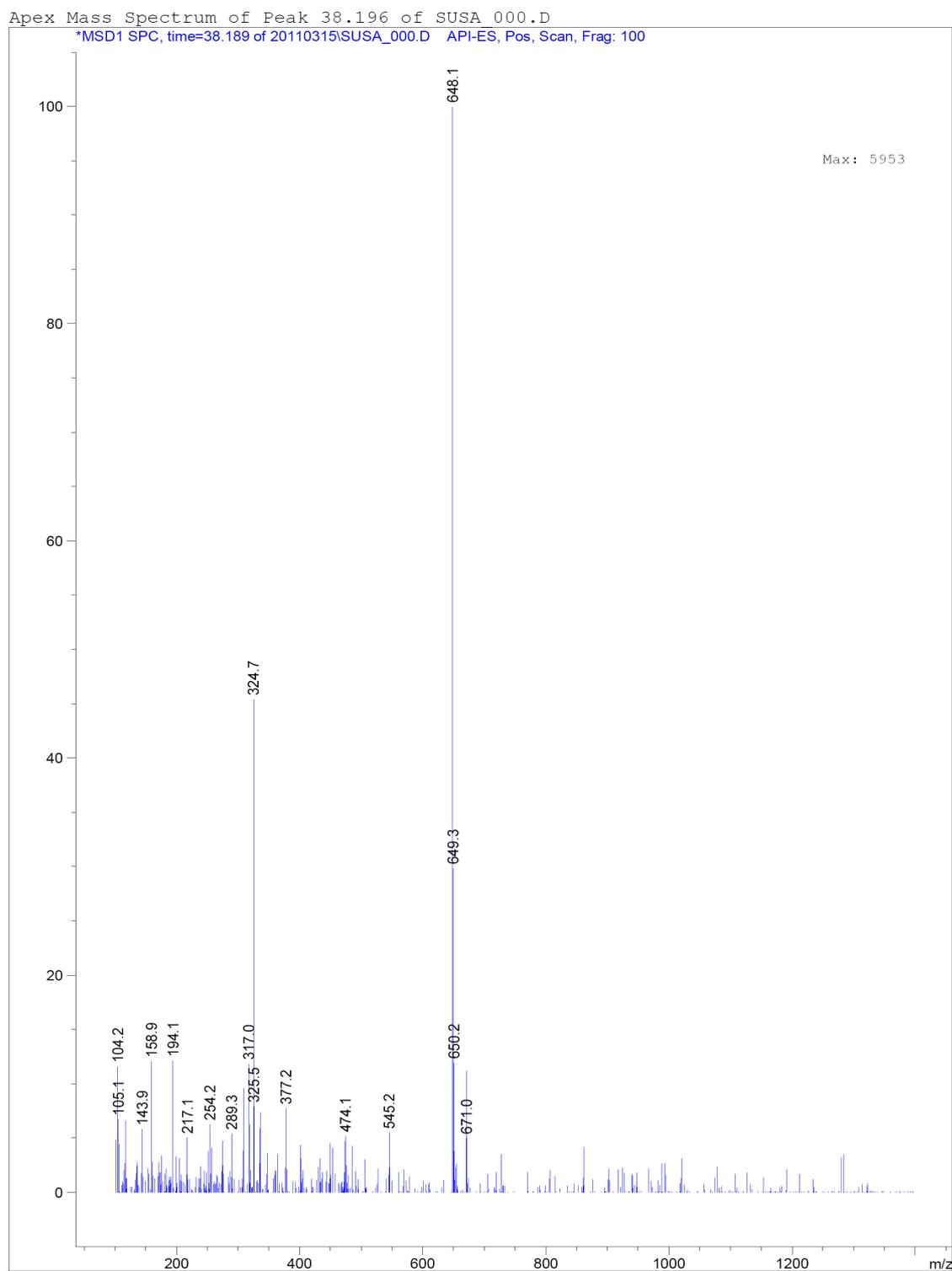


Figura 4.30: Espectro de masas obtenido para el pico número 10 (t_R : 38,4 minutos)

En este espectro se revela la presencia de Bz-Arg-(Ala)₅-OMe (igual a 648,1, 649,3 y 650,2), otro producto de polimerización del tipo Bz-Arg-(Ala)_n-OMe. También puede observarse la señal correspondiente a la especie con doble carga [m/z (M+2H⁺) igual a 324,7 y 325,5].

Debido a la presencia de múltiples señales secundarias, y a las cuales no ha sido posible atribuirles un origen concreto, este pico es claramente el resultado de una mezcla compleja de compuestos.

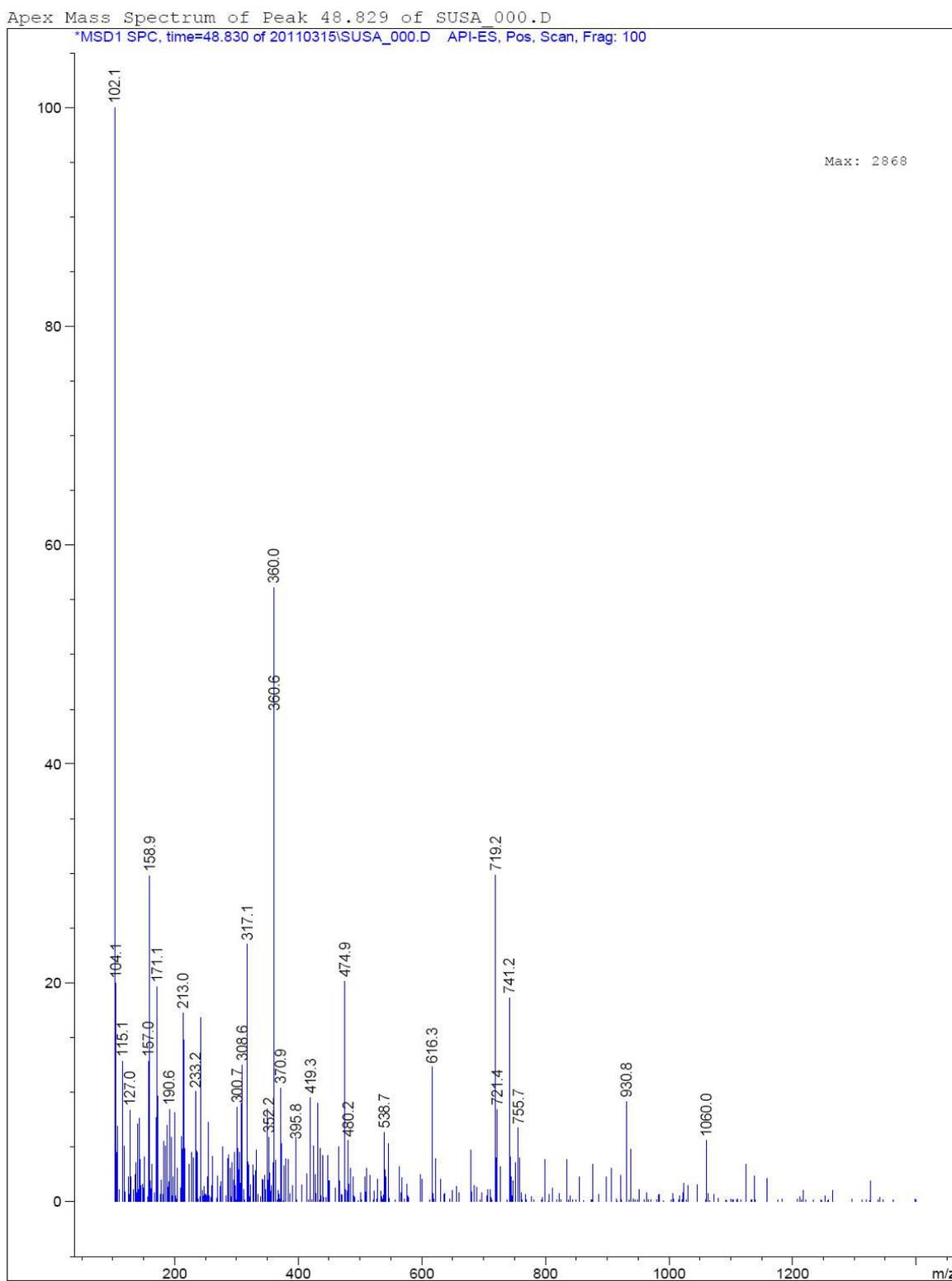


Figura 4.31: Espectro de masas obtenido para el pico número 11 (t_R : 48,83 minutos)

El espectro de masas para el último pico (Figura 4.31) observado en el cromatograma de la Figura 4.20, revela que el mismo también es resultado de una mezcla sumamente compleja de compuestos del tipo Bz-Arg-(Ala)_n-OMe y Bz-Arg-(Ala)_n-OH. Solo se ha podido identificar los siguientes productos:

- Bz-Arg-(Ala)₆-OMe; la especie M+H⁺ (m/z igual a 719,2 y 721,4) y M+2H⁺ (m/z igual a 360,0 y 360,6);
- Bz-Arg-(Ala)₉-OMe (M+H⁺) (m/z igual a 930,8);
- Bz-Arg-(Ala)₁₁-OH (M+H⁺) (m/z igual a 1060,0).

En resumen, la reacción entre Bz-Arg-OEt y Ala-OMe produjo, mediante catálisis enzimática, una mezcla compleja de péptidos de fórmula Bz-Arg-(Ala)_n-OMe y Bz-Arg-(Ala)_n-OH en el medio ACN:H₂O. Además se produjo una apreciable cantidad del producto de hidrólisis Bz-Arg-OH. La aparición de estos productos podría deberse a que los primeros derivados peptídicos formados actuarían como un acildonor y formarían el intermediario covalente acil-enzima con el sitio activo de papaína, para luego reaccionar con otra molécula del aminoácido metilado (Ala-OMe), resultando en los polímeros del tipo Bz-Arg-(Ala)_n-OMe, o bien ser deacilado por el agua dando el producto de hidrólisis. Esto ya fue observado en experiencias previas de reacciones catalizadas por papaína (Pera *et al.*, 1996). El mecanismo de la reacción propuesto se muestra en las figuras 4.32 y 4.33.

La aparición de esta clase de compuestos fue informada por primera vez para papaína por Sluyterman y Wijdenes (1972) quienes, al emplear como sustrato Leu-OMe para medir actividad porque papaína demostraba poca reactividad con Leu-NH₂, hallaron que luego de un período de inducción variable, aparecía un turbidez debida a un material insoluble, cuya identidad resultó ser polileucina. También se observó un comportamiento similar cuando se realizó la condensación enzimática entre Z-Ala y Phe-OMe (tanto por control cinético como por control termodinámico) catalizada por papaína y funastraína (Morcelle *et al.*, 2006). La correlación entre la formación de polímeros fuertemente hidrofóbicos y el desarrollo de turbidez en el medio puede indicar que tendría lugar la reacción de plasteínas, la que fuera definida como la conversión, en presencia de una peptidasa, de soluciones concentradas de péptidos solubles en mezclas de reacción

conteniendo una gran catidad de material insoluble tanto en medio acuoso como solventes orgánicos, aún después de tratamientos alcalinos o ácidos (Williams *et al.*, 2001).

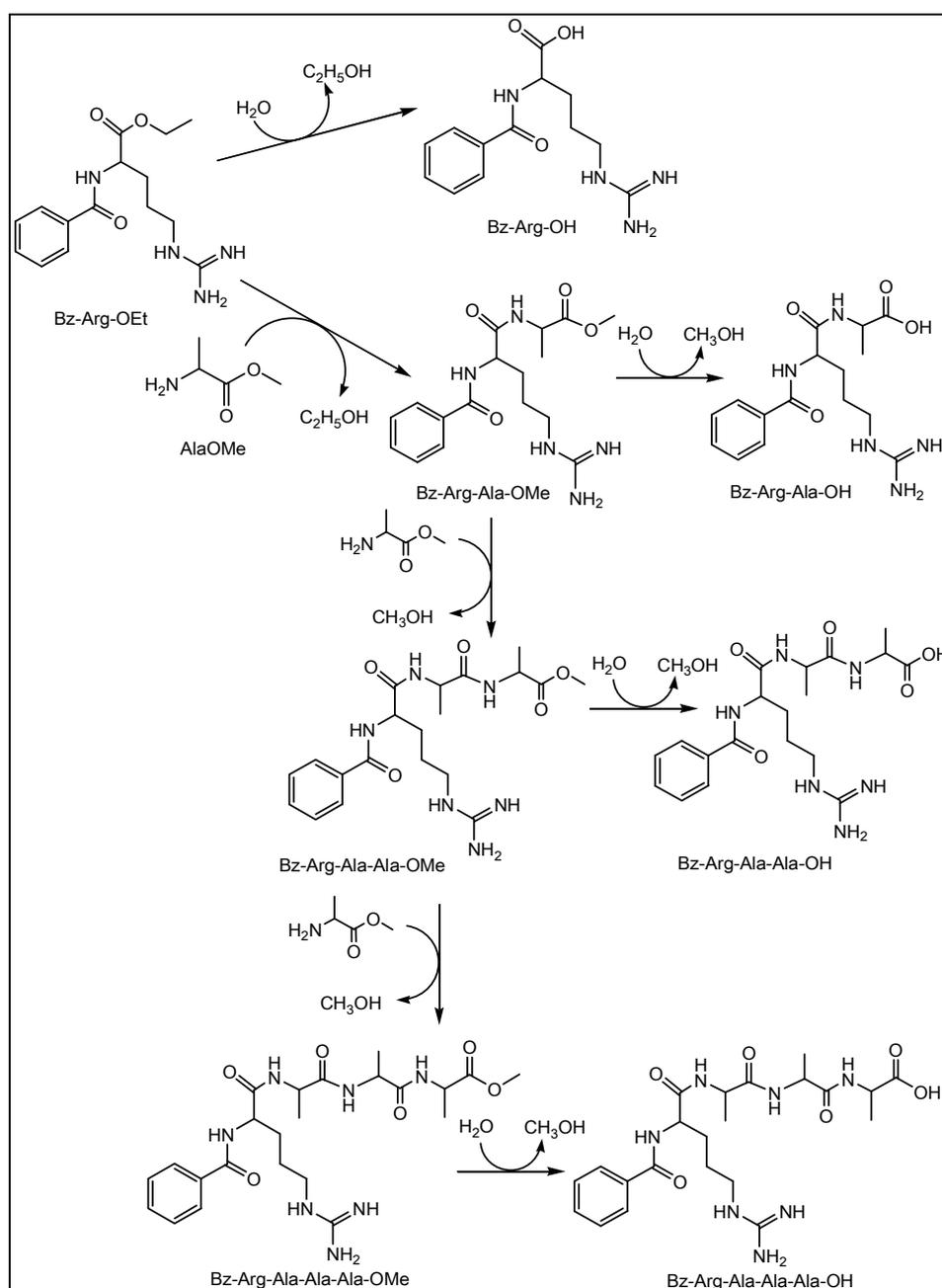


Figura 4.32: mecanismo de la reacción propuesto para la formación de productos del tipo Bz-Arg-(Ala)_n-OMe y del tipo Bz-Arg-(Ala)_n-OH

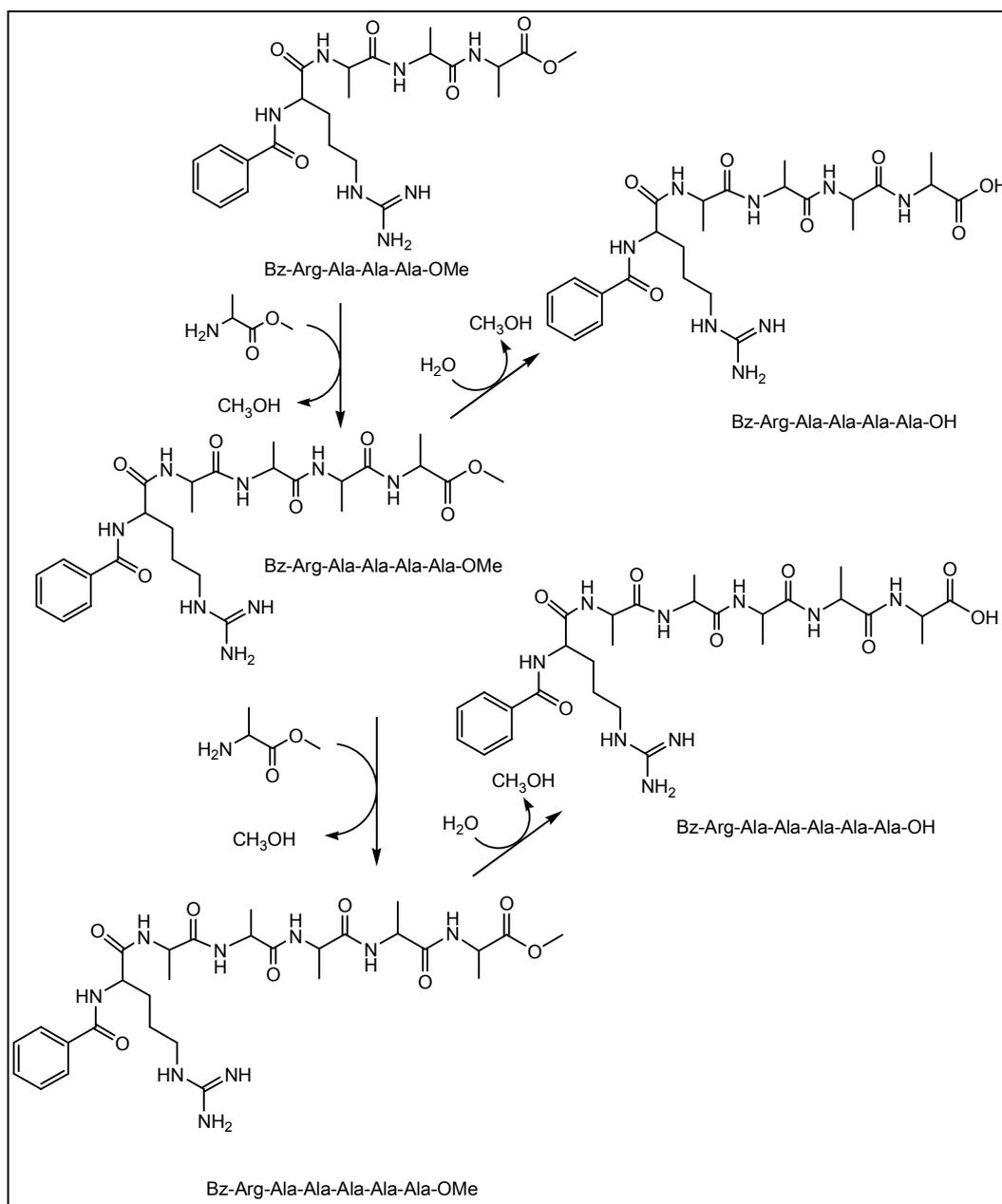


Figura 4.33: mecanismo de la reacción propuesto para la formación de productos del tipo Bz-Arg-(Ala)_n-OMe y del tipo Bz-Arg-(Ala)_n-OH

4.2. Aplicación de la metodología QSAR-3D para la obtención de modelos cualitativos de predicción de actividad sobre receptores IN de los derivados Bz-Arg

4.2.1. Modelado: Selección de un subconjunto de d descriptores

Se generaron 36 combinaciones al azar de los 20 bloques de descriptores calculados empleando el programa Dragon4.0[®]. Entre estos bloques encontramos, por ejemplo, descriptores constitucionales, descriptores topológicos, índices de conectividad, descriptores de geometría molecular, propiedades moleculares, entre otros.

La calidad de los 36 modelos obtenidos mediante regresión lineal múltiple y análisis discriminante (véase 3.2.3.) fue analizada y se seleccionaron aquellos que cumplieron los siguientes requisitos:

- ✓ entre el 85 y el 95% de buenas clasificaciones en el conjunto de entrenamiento
- ✓ λ de Wilk < 0,5

Así se seleccionaron 7 modelos (véase anexo 8.2.3.), denominados teniendo en cuenta los números de bloques de descriptores utilizados para generarlos (modelo 0314, modelo 0513, modelo 0715, modelo 1617, modelo 030515, modelo 031015 y modelo 091320).

4.2.2. Validación

4.2.4.1. Validación Externa

Se utilizaron 6 compuestos activos y 7 compuestos inactivos para evaluar la capacidad predictiva de los 7 modelos y se seleccionaron aquellos que mostraban entre el 85 y el 95% de buenas clasificaciones. Los parámetros estadísticos F y λ , y los % de buenas clasificaciones para el conjunto de entrenamiento y para el conjunto de prueba, para los 7 modelos se muestran en la tabla 4.6.

Tabla 4.6: Resultados de la validación externa: parámetros estadísticos de los modelos y sus % de buenas clasificaciones

Modelo	F	Valor p	λ	% aciertos en entrenamiento	% aciertos en prueba
0314	20,15061	0,000000	0,3732433	87,50	84,62%
0513	21,69914	0,000000	0,3560922	90,00	84,62%
0715	34,66426	0,000000	0,2571561	92,50	84,62%
1617	28,75016	0,000000	0,2944774	92,50	69,23%
030515	20,24125	0,000000	0,372194	87,50	92,31%
031015	20,33204	0,000001	0,4764108	85,00	84,62%
091320	20,90089	0,000000	0,3647318	90,00	76,92%

Los modelos 1617 y 091320 fueron descartados ya que el porcentaje de aciertos en el conjunto de prueba fue menor al 85%. Debido a la gran diferencia entre la capacidad predictiva de estos modelos dentro del conjunto de entrenamiento respecto a su capacidad de predicción dentro del conjunto de prueba, se consideró que son modelos sobre-ajustados.

A partir de los resultados de esta validación y teniendo en cuenta los valores de los estadísticos F y λ , puede concluirse que los mejores modelos son el modelo 030515 y el modelo 0715. Sin embargo los otros modelos (0314, 0513 y 031015) no fueron descartados y se sometieron también a validación interna.

4.2.4.2. Validación Interna

Los 5 modelos que superaron la validación externa fueron luego validados internamente mediante el ensayo de aleatorización de Fisher y validación cruzada Leave-Group-Out.

Los resultados del ensayo de aleatorización (tabla 4.7) se representan mediante el % de aciertos en el conjunto de entrenamiento para el modelo original, el % promedio (% de

aciertos en el conjunto de entrenamiento de los 40 modelos aleatorizados y el intervalo de confianza 99%.

Tabla 4.7: Resultados de la aleatorización de Fisher. El valor de significancia (α) utilizado para calcular la confianza fue 0.01.

	Modelo original	Modelos aleatorizados	
Modelo	% aciertos en entrenamiento	⟨%⟩ aciertos en entrenamiento	Intervalo de confianza
0314	87,50	76,81	76,81+/-2,14
0513	90,00	72,81	72,81+/-2,83
0715	92,50	73,56	73,56+/-2,51
030515	87,50	72,79	72,79+/-2,46
031015	85,00	72,38	72,38+/-2,21

Puede observarse que el % de aciertos en los modelos originales se encuentra siempre fuera del intervalo de confianza correspondiente a los modelos aleatorizados. Esto indica que los modelos obtenidos por correlaciones azarosas tienen un desempeño más pobre que el modelo original, lo que sugiere que el modelo original no ha sido obtenido por correlación azarosa entre la variable dependiente y los descriptores moleculares.

Por otro lado, los resultados de la validación Leave-5-Out se expresaron empleando el % de aciertos en el conjunto de prueba para cada modelo Leave-5-Out y el valor promedio (⟨%⟩) de los 16 modelos Leave-5-out. Los resultados obtenidos para cada modelo se muestran en la tabla 4.8.

Tabla 4.8 : Resultados de la validación LGO. El valor de significancia (α) utilizado para calcular la confianza fue 0.05.

	Modelo 030515		Modelo 0314		Modelo 0513		Modelo 031015		Modelo 0715	
	%entrenamiento	%prueba								
modelo original	87,5	92,31	87,5	84,62	90	88,89	87,5	84,62	92,5	84,62
modelos LGO	85,71	94,44	85,71	88,89	91,43	83,33	82,86	88,89	91,43	83,33
	88,57	88,88	88,57	77,78	91,43	88,89	82,86	88,89	94,29	83,33
	91,43	88,88	88,57	83,33	88,57	83,33	82,86	88,89	94,29	77,78
	82,86	94,44	91,43	77,78	85,71	88,89	82,86	88,89	94,29	88,89
	85,71	94,44	91,43	83,33	88,57	88,89	82,86	88,89	97,14	88,89
	88,57	77,78	88,57	88,89	88,57	88,89	88,57	77,78	91,43	77,78
	91,43	88,88	88,57	77,78	91,43	83,33	88,57	77,78	94,29	88,89
	88,57	88,88	94,29	66,67	94,29	83,33	85,71	83,33	94,29	77,78
	91,43	83,33	88,57	77,78	91,43	83,33	85,71	88,89	94,29	83,33
	91,43	88,88	91,43	66,67	91,43	83,33	88,57	77,78	94,29	83,33
	88,57	88,88	88,57	72,22	88,57	83,33	85,71	88,89	94,29	83,33
	88,57	88,88	91,43	77,78	94,29	83,33	88,57	83,33	91,43	88,89
	91,43	88,88	91,43	77,78	94,29	77,78	88,57	77,78	100,00	83,33
	85,71	66,67	85,71	83,33	91,43	88,89	85,71	83,33	94,29	72,22
	85,71	94,44	85,71	88,89	88,57	83,33	82,86	88,89	91,43	83,33
85,71	83,33	94,29	83,33	91,43	83,33	84,21	94,44	97,14	88,89	
Promedios LGO	88,21	87,49	89,64	79,51	90,71	84,72	85,44	85,42	94,29	83,33

Los modelos 0314, 031015 y 0715 muestran más de dos modelos LGO con porcentaje mucho menor al valor de corte aceptado (85%). Estos modelos LGO están fuertemente sustentados por aquellas estructuras que fueron excluidas durante el proceso de validación cruzada, por lo que presentan bajos valores predictivos al remover dichos compuestos, indicando que estos tres modelos son poco robustos. Con este mismo criterio, el modelo más robusto parece ser el modelo 0513, ya que solo uno de los modelos LGO generados durante la validación muestra un % de aciertos en el conjunto de prueba mucho menor al porcentaje aceptado. Este modelo fue elegido para realizar la predicción de actividad de los productos sintetizados. En la tabla 4.9 se muestra el modelo seleccionado y sus parámetros estadísticos.

Tabla 4.9: Modelo 0513; resumen de los valores de los coeficientes de regresión y valores de los parámetros estadísticos.

Modelo 0513	Beta	B	Valor-p
Constante		-3,55071	0,123340
RDF030p	0,414194	0,24114	0,000570
RDF020m	-0,450087	-3,61875	0,000181
IVDE	0,256944	2,97138	0,015764
F = 21,69914 p = 0,000000 Lambda de Wilk: 0,3560922			

4,2,3, Análisis del modelo seleccionado

En el modelo 0513, el valor de la variable dependiente df depende del valor de los descriptores IVDE, RDF020m y RDF030p,

Los coeficientes de regresión normalizados (Beta) muestran que hay una diferencia marcada en las contribuciones de los diferentes descriptores al valor final de la fd , Los descriptores RDF contribuyen en magnitud semejante, sin embargo esta contribución es en sentidos opuestos, Por otro lado, el descriptor IVDE, contribuye en forma positiva al valor

de df , aunque su contribución relativa es (en magnitud) alrededor de la mitad de la contribución de los descriptores RDF,

Los descriptores RDF (Radial Distribution Function) son descriptores 3D que se calculan a partir de la distribución de probabilidad de encontrar un átomo en un volumen esférico de cierto radio r , empleando diferentes propiedades atómicas (polarizabilidad, electronegatividad, masa atómica) para caracterizar los tipos de átomos, Por ejemplo, RDF030p considera esferas de radio r igual a 3,0 armstrongs y polarizabilidades atómicas, en tanto RDF020m toma en esferas de radio 2,0 armstrong y tiene en cuenta las masas atómicas, Pueden encontrarse más detalles al respecto en el trabajo de Hemmer *et al.*, 1999, La función de distribución de probabilidad utilizada es:

$$g(r) = f \sum_{i=1}^{N-2} \sum_{j>i}^N A_i A_j e^{-B(r-r_{ij})^2}$$

Donde f es un factor escalar, N es el número de átomos de la molécula, A denota la propiedad atómica de los átomos i y j , B es un parámetro de ventana (en principio, inversamente proporcional a la raíz de la longitud de los intervalos Δr utilizados para calcular g), y r_{ij} es la distancia entre los átomos i y j ,

Como podemos observar en la expresión, para dos átomos i y j separados entre si por una distancia similar a r (el valor del radio de la esfera tomada por el descriptor), el término de la sumatoria correspondiente a dichos átomos tomará mayores valores si los mismos tienen valores relativamente altos de la propiedad atómica A , En particular, con respecto al descriptor RDF030p, y teniendo en cuenta que el coeficiente de regresión que acompaña a este descriptor es positivo, se espera entonces que aquellas moléculas que son activas presenten mayor cantidad de pares atómicos con átomos de alta polarizabilidad separados por distancias cercanas a 3 amstromgs, Esto se verifica para algunos de los compuestos activos del conjunto de entrenamiento, donde hay una presencia importante de

anillos aromáticos y anillos fusionados, donde las distancias atómicas entre los carbonos que forman parte del mismo anillo y entre carbonos de anillos vecinos, se encuentra alrededor de los 3 armstrong,

Inversamente, el descriptor RDF020m esta acompañado de un coeficiente negativo, por lo que la presencia de pares de átomos que incluyen uno o dos heteroátomos separados por distancias cercanas a 2 armstrongs en una determinada estructura, contribuiría en forma negativa al valor de df, En particular, podemos observar que en las estructuras inactivas del conjunto de entrenamiento aparecen zonas de la molécula con gran densidad de heteroátomos (3 a 5 heteroátomos) muy próximos entre sí (obsérvese la abundancia de grupos guanidonio, tioguanidonio y urea, entre otros),

Por otro lado, el contenido medio de información sobre la equivalencia de grados de los vértices IVDE es un descriptor topológico y es un índice de información calculado a partir de la entropía de Shannon (contenido de información), que se computa según la siguiente expresión (Melagraki *et al.*, 2006):

$$IVDE = - \sum_{g=1}^G \frac{g_F}{N} \log_2 \frac{g_F}{N}$$

donde g_F es la frecuencia de vértices de grado g y N es el número de átomos, G es el valor máximo de orden de vértice para la molécula, Así, valores bajos de este descriptor indican o bien una estructura muy lineal o bien una muy ramificada, y valores mayores indican diferentes niveles de ramificación,

Para continuar con el análisis del modelo 0513, se construyó un Diagrama de Distribución Farmacológica, para los compuestos del conjunto de entrenamiento (véase 3,2,5,), Los resultados obtenidos se muestran en la figura 4,34,

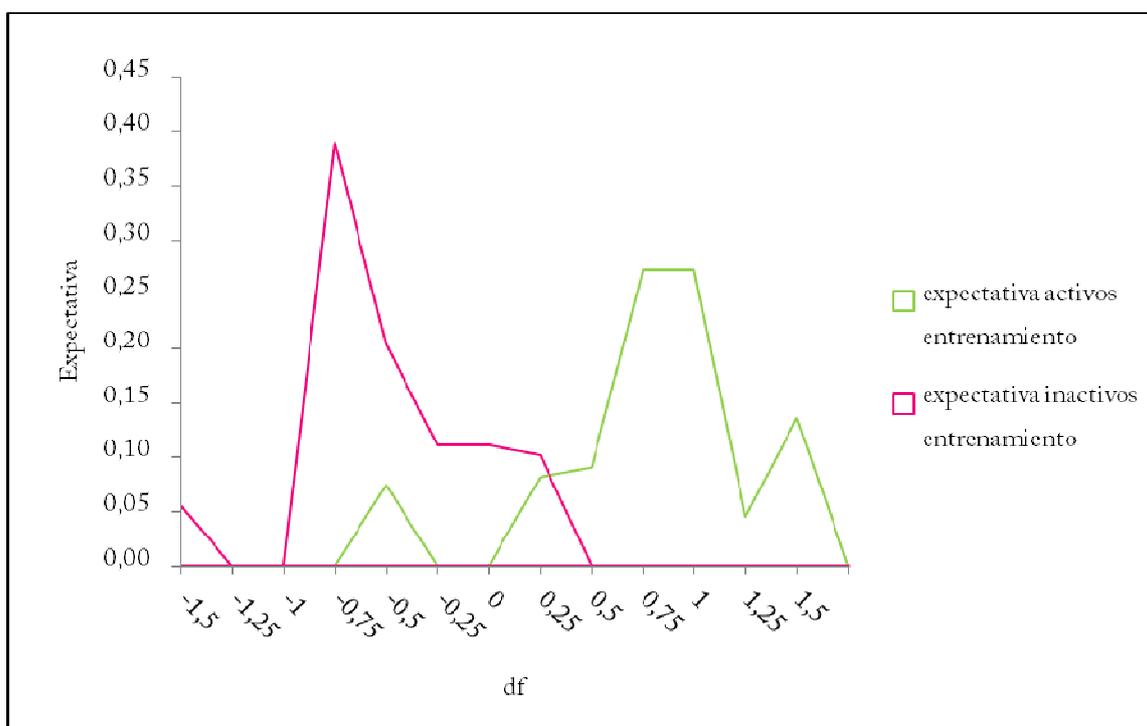


Figura 4,34: Diagrama de distribución farmacológica para los compuestos del conjunto de entrenamiento del modelo 0513,

En esta gráfica de áreas puede verse que parte del área correspondiente a los compuestos inactivos se superpone con el área correspondiente a los activos, Este resultado refleja que la capacidad discriminante de la función no es perfecta, Recordemos que el λ de Wilk de este modelo es 0,356 y, aunque se considera un valor aceptable, no denota una capacidad discriminante ideal, Aún así, la probabilidad de encontrar un compuesto inactivo en la zona de los activos es baja, para valores de df mayores a 0,5, En consecuencia, en la posterior utilización de este modelo, y con el objetivo de disminuir el número de falsos positivos (compuestos que se envíen a ensayo biológico sin resultados positivos) a costa de aumentar el número de falsos negativos (compuestos descartados a pesar de ser activos), deberán seleccionarse aquellos compuestos con valores de df entre 0,5 y 1,75, correspondientes a la zona de mínima superposición entre las dos categorías consideradas para el conjunto de entrenamiento,

4,2,4, Aplicación del modelo obtenido para la predicción de actividad sobre receptores IN de los derivados Bz-Arg

Finalmente se utilizó el modelo seleccionado para predecir el valor de df de los productos de las reacciones de síntesis ensayadas, Los resultados se muestran en la tabla 4,10,

Tabla 4,10: Resultados de la predicción de actividad de los productos de síntesis

Producto	RDF020m	RDF030p	IVDE	df
Bz-Arg-Gliol	1,111	2,94	1,76	-1,63256085
Bz-Arg-Gli	0,928	2,881	1,786	-0,90730098
Bz-Arg-Ala-OMe	0,977	5,174	1,852	-0,33557463

El modelo 0513 predice que todos estos productos son inactivos, El hecho de que los tres productos muestren valores similares para cada descriptor era esperable dada la similitud estructural y química de los mismos (figura 4,35), Sin embargo el producto de síntesis que contiene alanina en su estructura muestra un valor de RDF030p significativamente superior al mostrado por los otros dos productos, Esto podría tener que ver con la presencia de los dos grupos metilo adicionales en esta estructura,

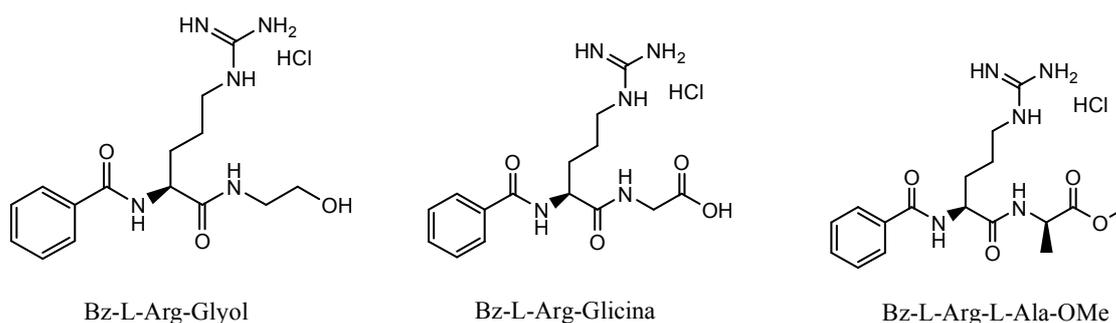


Figura 4,35: Estructuras de los productos de síntesis

En cuanto al valor del descriptor IVDE, todos los productos muestran valores semejantes y esto reflejaría el hecho de que estas las estructuras tienen un nivel similar de ramificación en su estructura,

Por otro lado si se compara el valor que los descriptores toman para los productos, con los valores que estos mismos muestran para los compuestos del conjunto de entrenamiento (tabla 4,11), puede observarse que tanto el valor del descriptor IVDE como el del descriptor RDF020m son muy similares a los observados en general para todos los compuestos del conjunto de prueba, Sin embargo, el valor que toma el descriptor RDF030p para los productos es mucho menor al valor que típicamente muestran los compuestos inactivos del conjunto de prueba, Esto nos llevaría a pensar que éste es el descriptor que está influenciando principalmente al valor de la *df*, y por tanto, en la actividad de los productos,

Tabla 4,11: Valores que presentan los descriptores para algunos de los compuestos del conjunto de entrenamiento, La letra A o I que antecede al nombre de la estructura indica que la misma se trata de un compuesto activo o inactivo, respectivamente,

Compuesto	RDF020m	RDF030p	IVDE	valor predicho
I1013aconf2	1,465	7,03	1,728	-2,0224199
I927conf2	1,101	6,442	1,677	-0,9985256
I67conf5	1,131	7,301	1,681	-0,8880633
I924aconf3	0,981	5,112	1,701	-0,8136787
I928conf5	1,101	7,438	1,666	-0,7910354
I610conf3	0,942	7,257	1,491	-0,7792919
I1018bconf5	1,1	6,488	1,752	-0,7609609
I69conf3	0,928	7,162	1,491	-0,7515377
I918aconf1	1,047	6,473	1,701	-0,7243247
A23Rconf3	0,88	8,66	1,848	0,8441726
A313conf5	0,981	12,618	1,655	0,8596347
A317conf3	0,797	9,561	1,693	0,9012321
A713conf1	0,822	8,921	1,782	0,9208866
A318conf1	0,839	10,354	1,693	0,9404687
A717conf2	0,897	9,504	1,846	0,9802333
A7Bconf5	0,821	10,136	1,768	1,1758911
A23Sconf5	0,992	12,047	1,848	1,2556138
A315conf4	0,826	12,116	1,655	1,2994886
A22conf4	0,969	12,095	1,851	1,3593339

Dentro de los posibles avances en la interpretación y análisis del modelo se propone realizar un análisis mas detallado de las estructuras de los compuestos que forman parte del conjunto de entrenamiento, con el objetivo de evidenciar aquellos patrones estructurales que justifiquen los valores que estos descriptores adoptan tanto en las estructuras activas como en las estructuras inactivas, particularmente en el caso del descriptor RDF030p, Esto permitirá avanzar en la interpretación fisicoquímica del descriptor e hipotetizar posibles modificaciones que podrían eventualmente realizarse a estos compuestos, u a otros, con el objetivo de modificar su nivel de actividad sobre receptores del tipo integrinas, En otras palabras, si logramos asociar ciertos patrones estructurales a valores elevados de un descriptor u otro, podremos pensar cuáles deberían ser las modificaciones que transformen a nuestros productos (inactivos) en compuestos activos,

5. Conclusiones

5.1. Síntesis enzimática de derivados de N^α -benzoil arginina (Bz-Arg)

- ~ Araujiaína presenta actividad proteolítica, incluso luego de un período breve de exposición a medios orgánicos (ACN, EtOH y MeOH).
- ~ Araujiaína presenta actividad luego de ser inmovilizada por adsorción en poliamida.
- ~ No hay síntesis de Bz-Arg-Glyol en la reacción efectuada utilizando Bz-Arg-OEt y glicinol, en ACN 0,25% (v/v) en H₂O, empleando araujiaína inmovilizada, aunque sí pudo observarse la formación del producto de hidrólisis del sustrato (Bz-Arg-OH)
- ~ No hay síntesis de Bz-Arg-Glyol en la reacción efectuada utilizando Bz-Arg-OEt y glicinol, en las mezclas ACN:H₂O y EtOH:H₂O (85:15), empleando araujiaína libre.
- ~ No hay síntesis de Bz-Arg-Gly en la reacción efectuada utilizando Bz-Arg-OEt y glicina, en la mezcla ACN: H₂O (85:15), empleando araujiaína libre.
- ~ La reacción entre Bz-Arg-OEt y Ala-OMe en las mezclas ACN:H₂O y EtOH:H₂O (85:15), empleando araujiaína libre como biocatalizador, produce una mezcla compleja de péptidos de fórmula Bz-Arg-(Ala)_n-OMe y Bz-Arg-(Ala)_n-OH, además de una cantidad apreciable del producto de hidrólisis Bz-Arg-OH.

5.2. Aplicación de la metodología QSAR-3D para la obtención de modelos cualitativos de predicción de actividad sobre receptores IN de los derivados Bz-Arg

- ~ Mediante la aplicación del análisis lineal discriminante se generaron 7 modelos clasificatorios que presentaron un porcentaje de buenas clasificaciones superior al 85%. Sin embargo, solo uno de ellos (el modelo 0513) superó en forma exitosa todas las etapas de validación, indicando robustez y buena capacidad predictiva.
- ~ A partir del diagrama de distribución farmacológica del modelo seleccionado, se observa que la probabilidad de encontrar falsos positivos se reduce para valores de df superiores a 0,5.
- ~ La implementación de este modelo al análisis de otros RGDmimetics permitirá el uso racional de recursos, ya que brindará, a priori, un criterio de selección de productos al momento de proyectar el trabajo de síntesis.

6. Proyecciones

El presente trabajo constituye solo la etapa inicial de un trabajo más profundo que podría incorporar:

- ~ La optimización del método de separación de los productos de síntesis por HPLC, lo que permitiría obtener fracciones más puras, y mejoraría el análisis analítico de los mismos.
- ~ La implementación de protocolos de síntesis modificados. Concretamente se propone incrementar las cantidades de TEA y enzima empleadas. Alternativamente al uso de mezclas conteniendo solventes orgánicos, podría ensayarse un medio 100% acuoso.
- ~ En este trabajo, todas las reacciones de síntesis se realizaron por control cinético. Como alternativa, podría ensayarse la síntesis por control termodinámico.
- ~ La optimización de las condiciones de reacción para mejorar los rendimientos de la síntesis de derivados conteniendo alanina.
- ~ La utilización de araujaína inmovilizada en poliamida u otros soportes, como catalizador para la síntesis de derivados de Bz-Arg-Ala.
- ~ La realización del ensayo biológico de actividad del producto sintetizado (Bz-Arg-Ala-OMe), y de otros productos que previamente sean analizados mediante el modelo QSAR generado, lo que constituiría la validación definitiva del mismo.
- ~ Optimización del modelo en si mismo aumentando el número y diversidad de compuestos utilizados como conjunto de entrenamiento. Además, el incremento en el número de estructuras utilizadas para la validación externa permitiría una mejor evaluación de la capacidad predictiva del mismo.

- ~ Realizar un análisis mas detallado de las estructuras de los compuestos que forman parte del conjunto de entrenamiento, con el objetivo de realizar una interpretación fisico-química de los descriptores incluidos en el modelo. Es decir, encontrar una relación entre los grupos presentes en las estructuras y los valores mostrados por los descriptores del modelo para estas mismas estructuras. Esto permitira sugerir modificaciones que podrían transformar compuestos inactivos en activos.
- ~ La implementación del modelo para la selección de posibles productos a sintetizar permitirá generar un criterio de prioridad, con la consecuente racionalización de los recursos humanos y económicos.

8. Anexos

8.1. Síntesis enzimática de derivados de N^α -benzoil arginina (Bz-Arg)

8.1.1. Cromatogramas obtenidos para los patrones de sustrato (Bz-Arg-OEt) y de sustrato hidrolizado (Bz-Arg-OH).

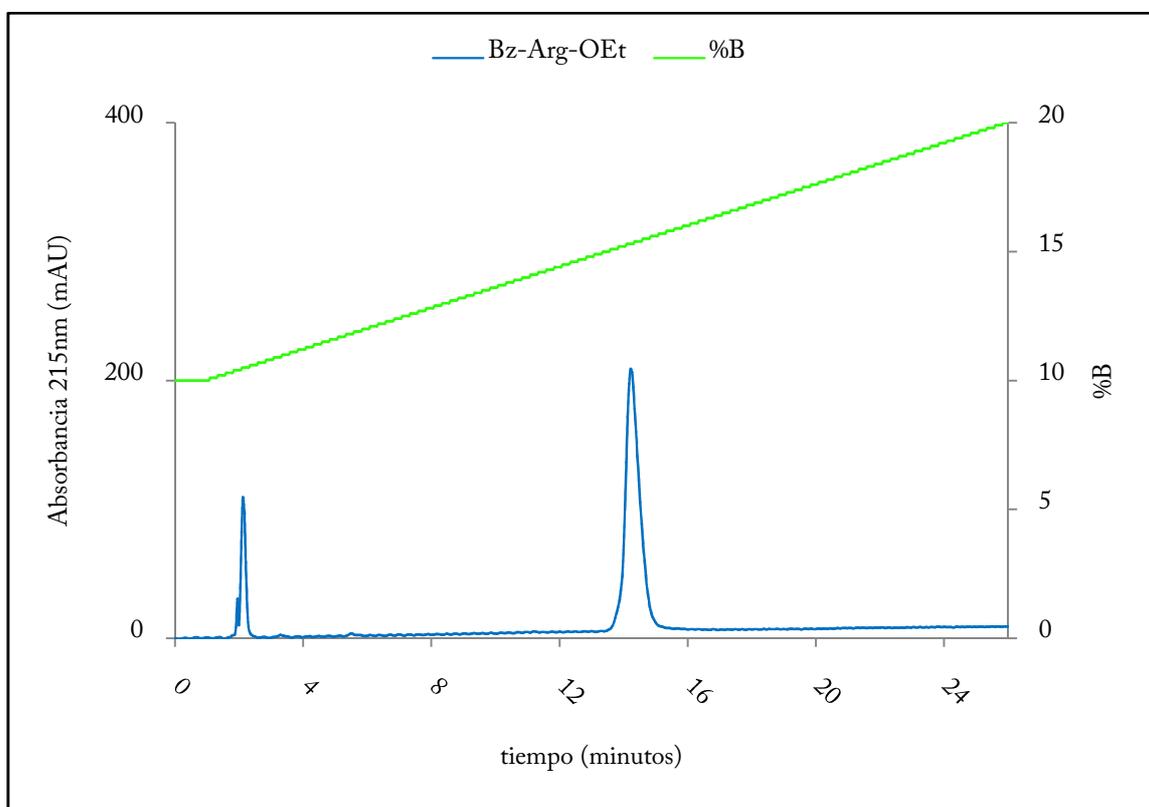


Figura 8.1: Cromatograma correspondiente al patron de sustrato (Bz-Arg-OEt).

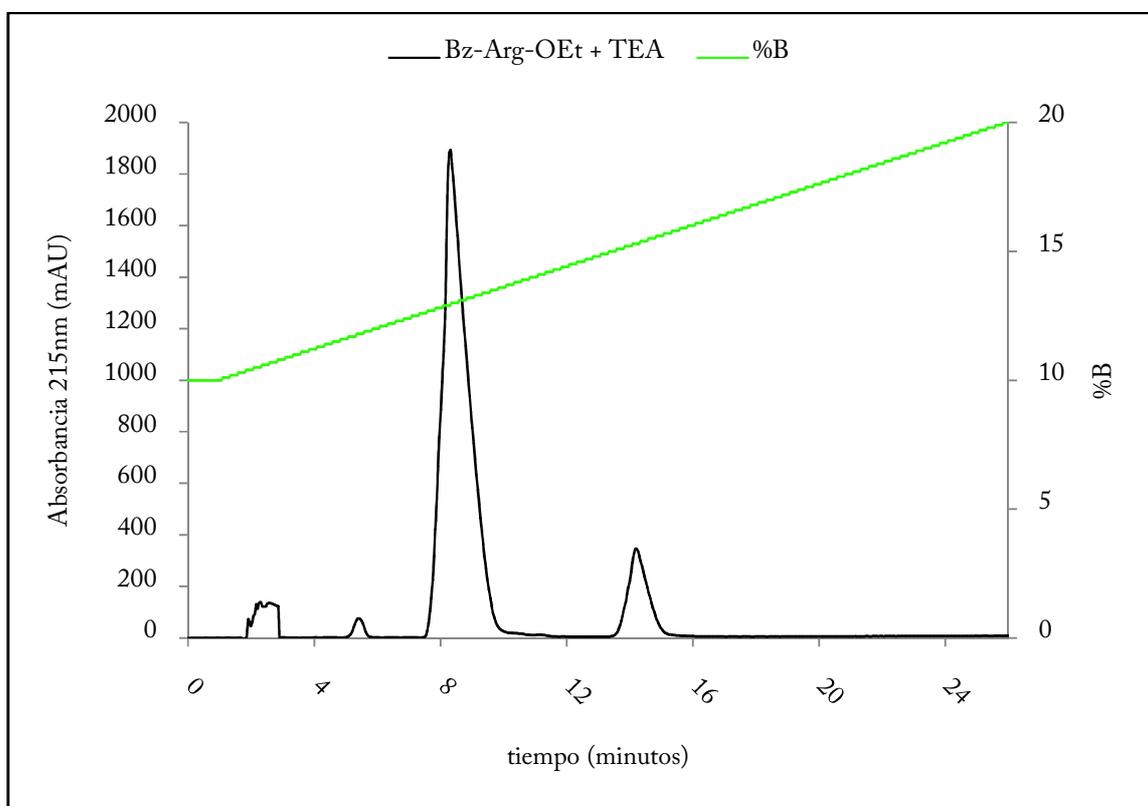


Figura 8.2: Cromatograma correspondiente al patrón de sustrato (Bz-Arg-OEt) conteniendo trietilamina (TEA) en cantidad equimolar.

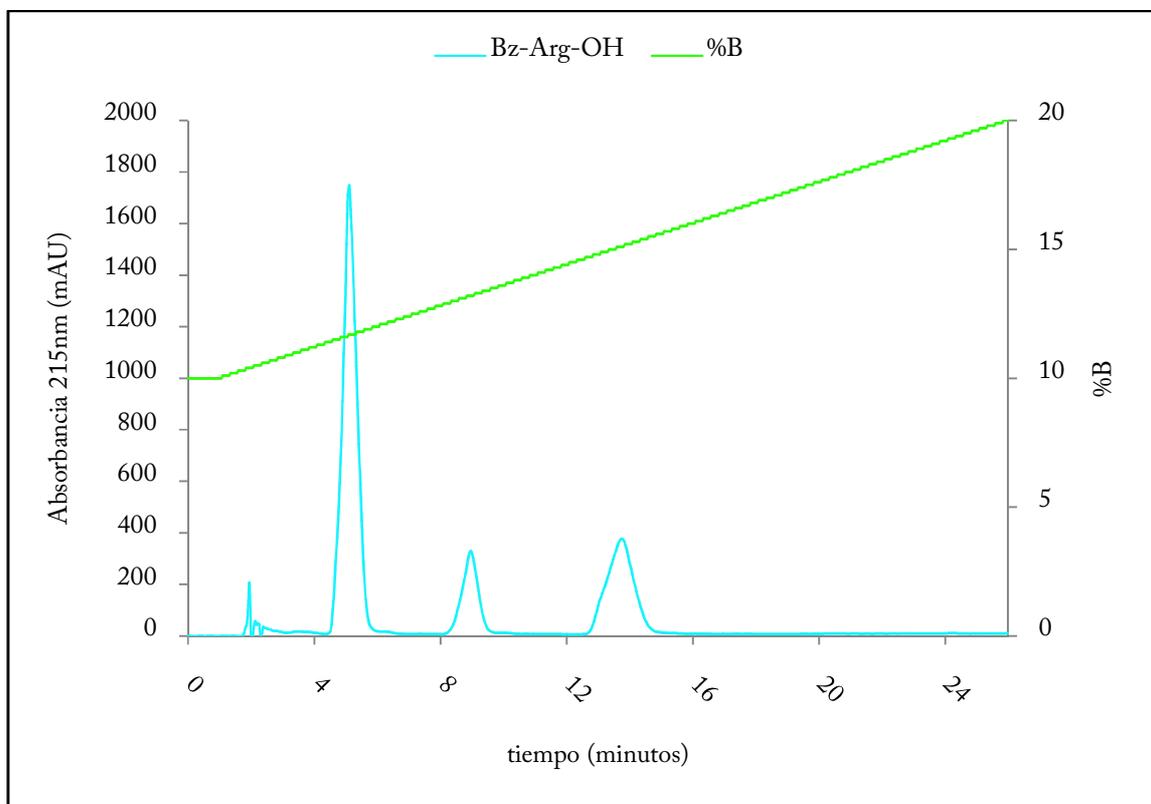


Figura 8.3: Cromatograma correspondiente al patron del sustrato hidrolizado (Bz-Arg-OH).

8.1.2. Cromatogramas obtenidos en la síntesis enzimática de derivados de Bz-Arg-Glyol utilizando glicinol como nucleófilo.

8.1.2.1. Primer protocolo de reacción: enzima inmovilizada y medio de bajo contenido acuoso

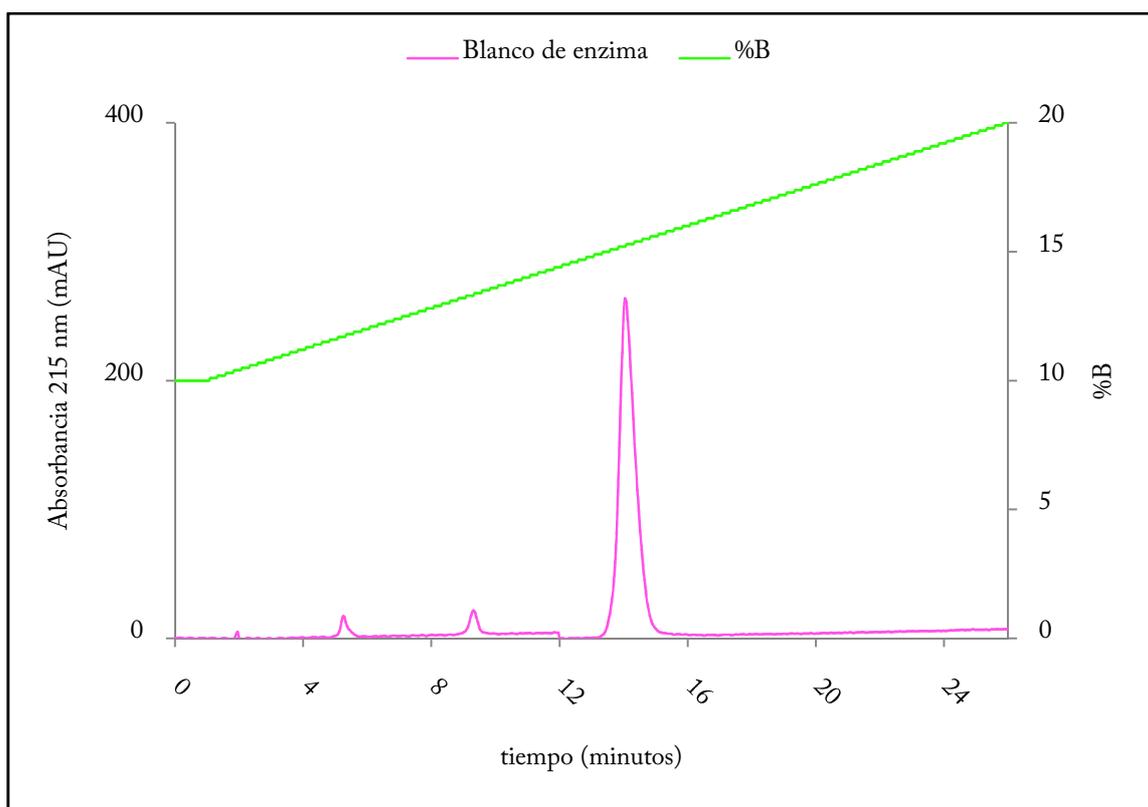


Figura 8.4: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en ACN 0,25% H₂O. Cromatograma correspondiente al blanco de enzima luego de 3 horas de incubación.

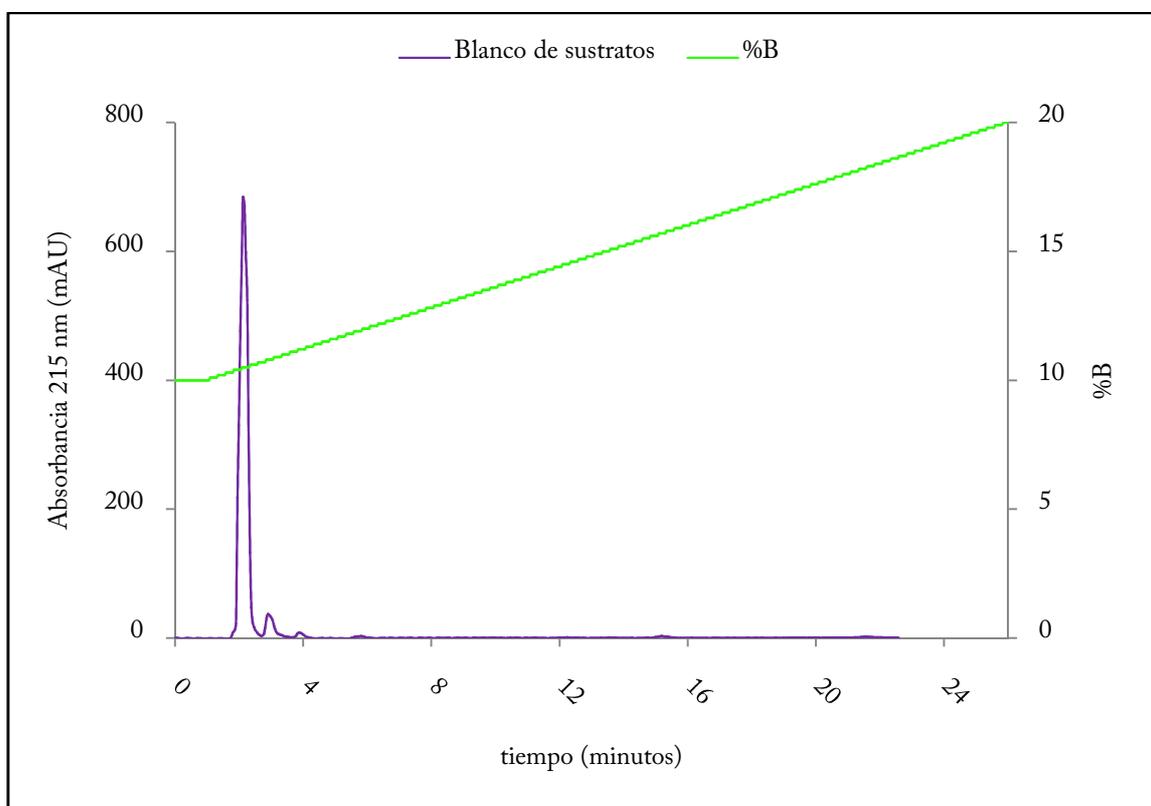


Figura 8.5: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en ACN 0,25% H₂O. Cromatograma correspondiente al blanco de sustratos de la reacción.

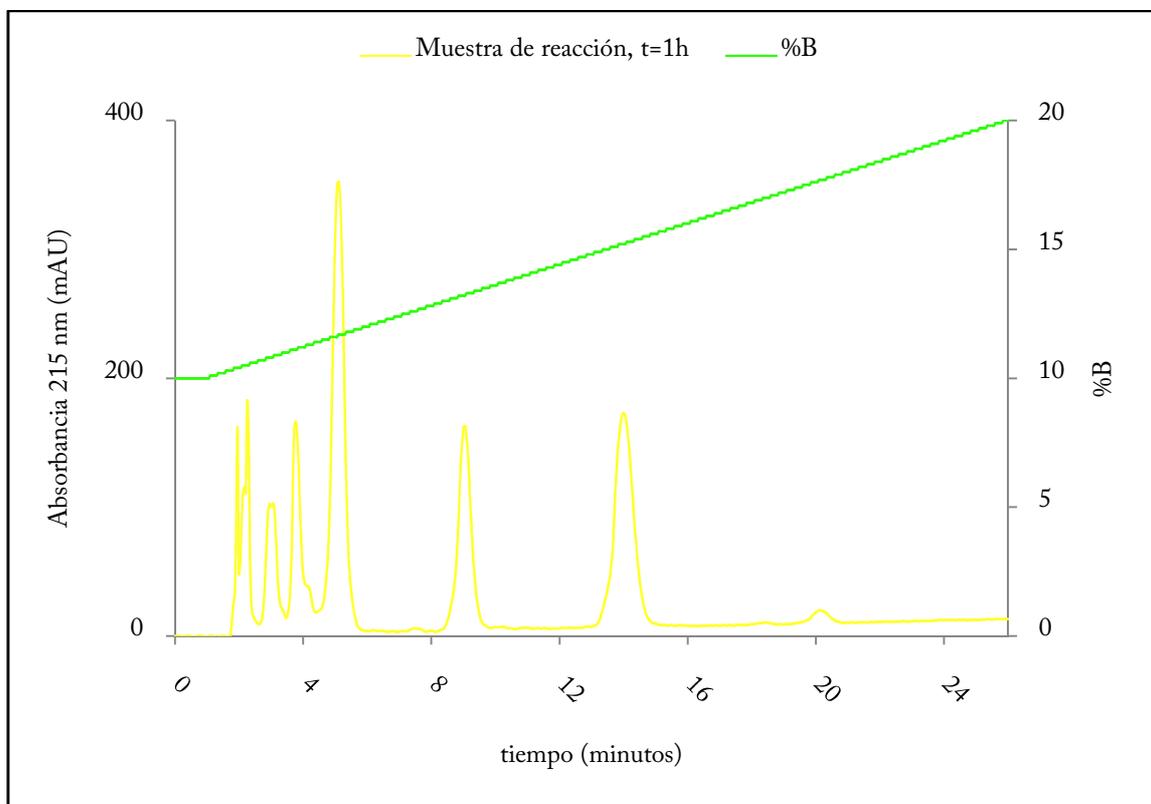


Figura 8.6: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en ACN 0,25% H₂O. Cromatograma correspondiente a una muestra de reacción tomada luego de 1hora de incubación.

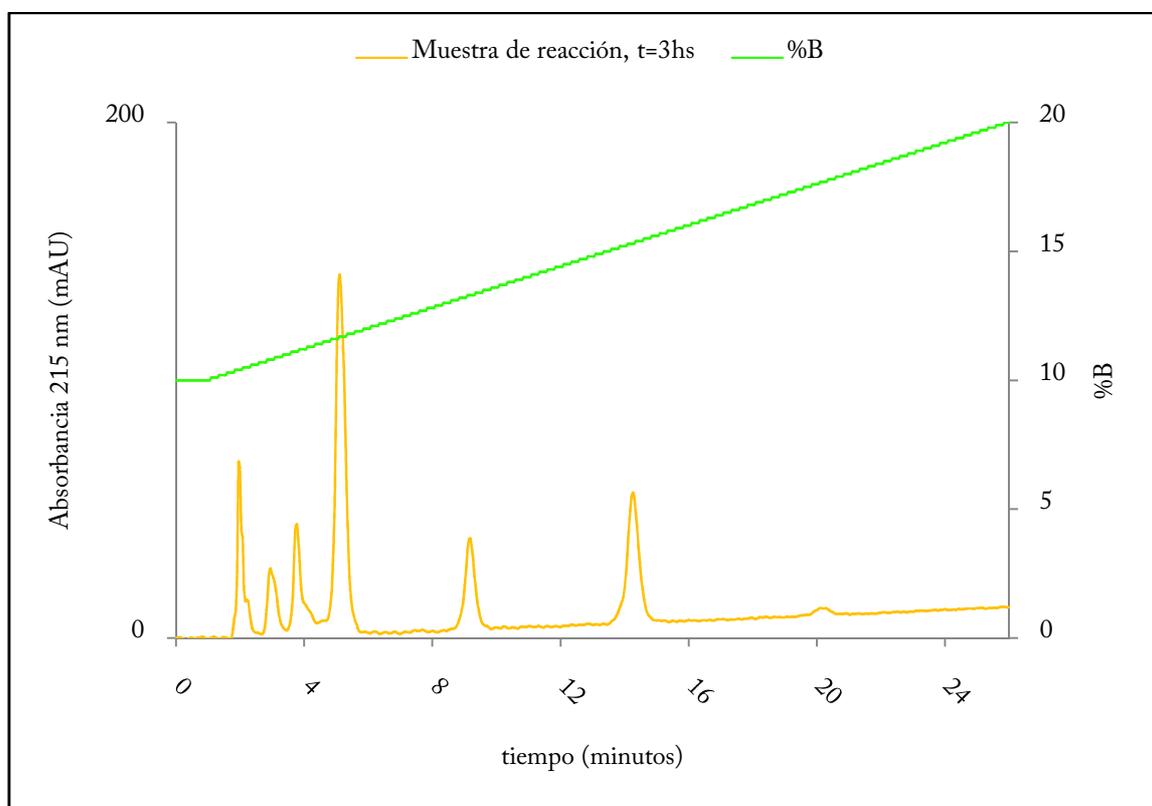


Figura 8.7: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en ACN 0,25% H₂O. Cromatograma correspondiente a una muestra de reacción tomada luego de 3 horas de incubación.

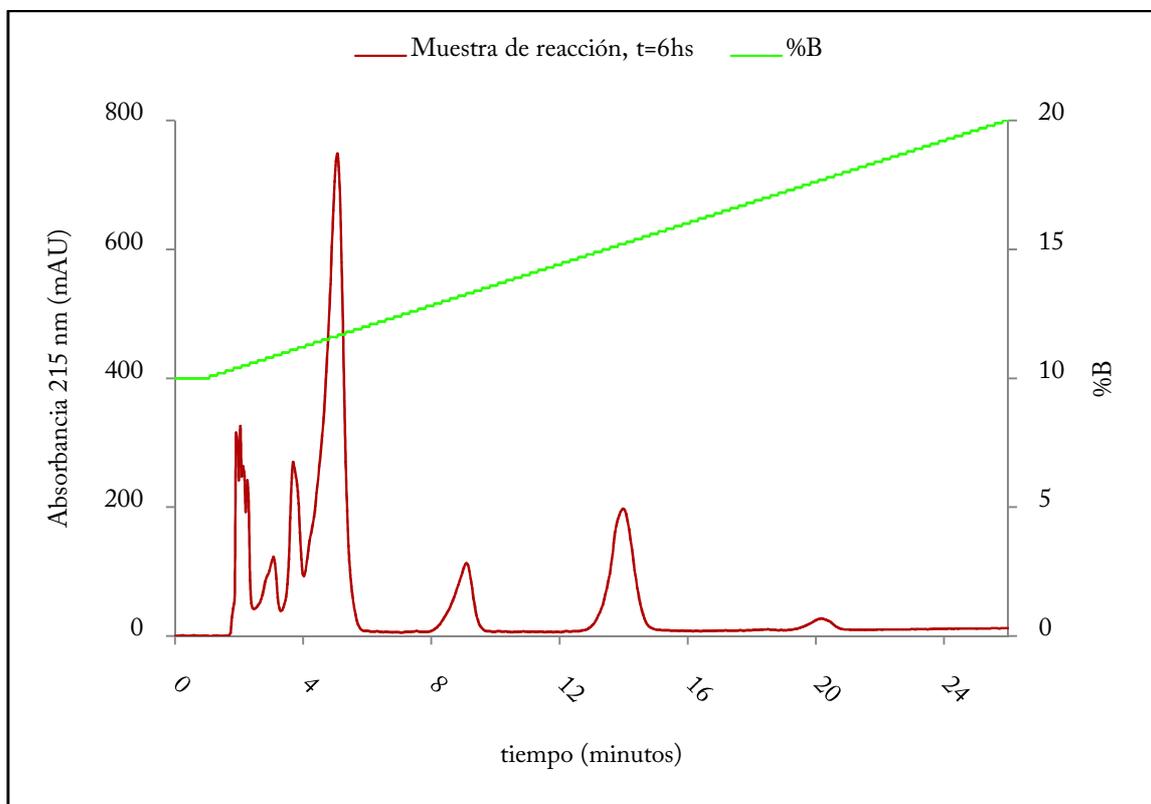


Figura 8.8: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en ACN 0,25% H₂O. Cromatograma correspondiente a una muestra de reacción tomada luego de 6 horas de incubación.

8.1.2.2. Segundo protocolo de reacción: enzima libre y medio de mediano contenido acuoso

Medio ACN:H₂O (85:15)

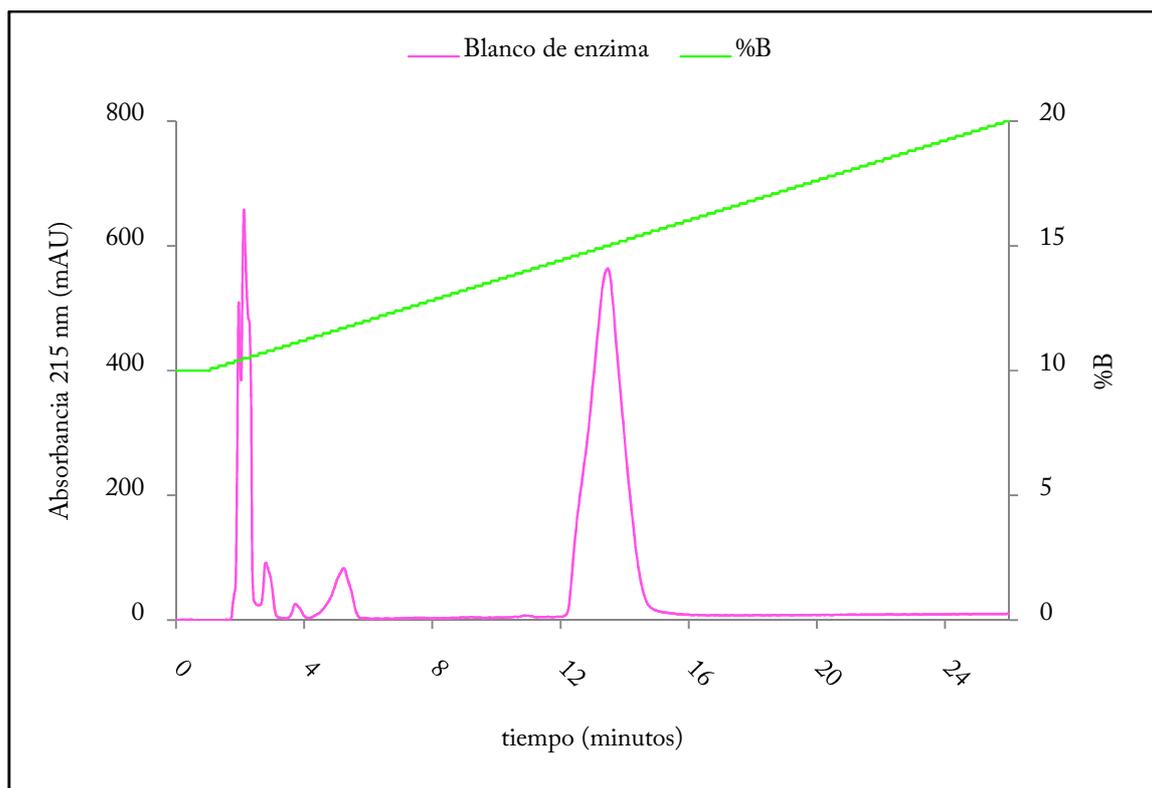


Figura 8.9: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en la mezcla ACN:H₂O (85:15). Cromatogramas correspondiente al blanco de enzima luego de 3 horas de reacción.

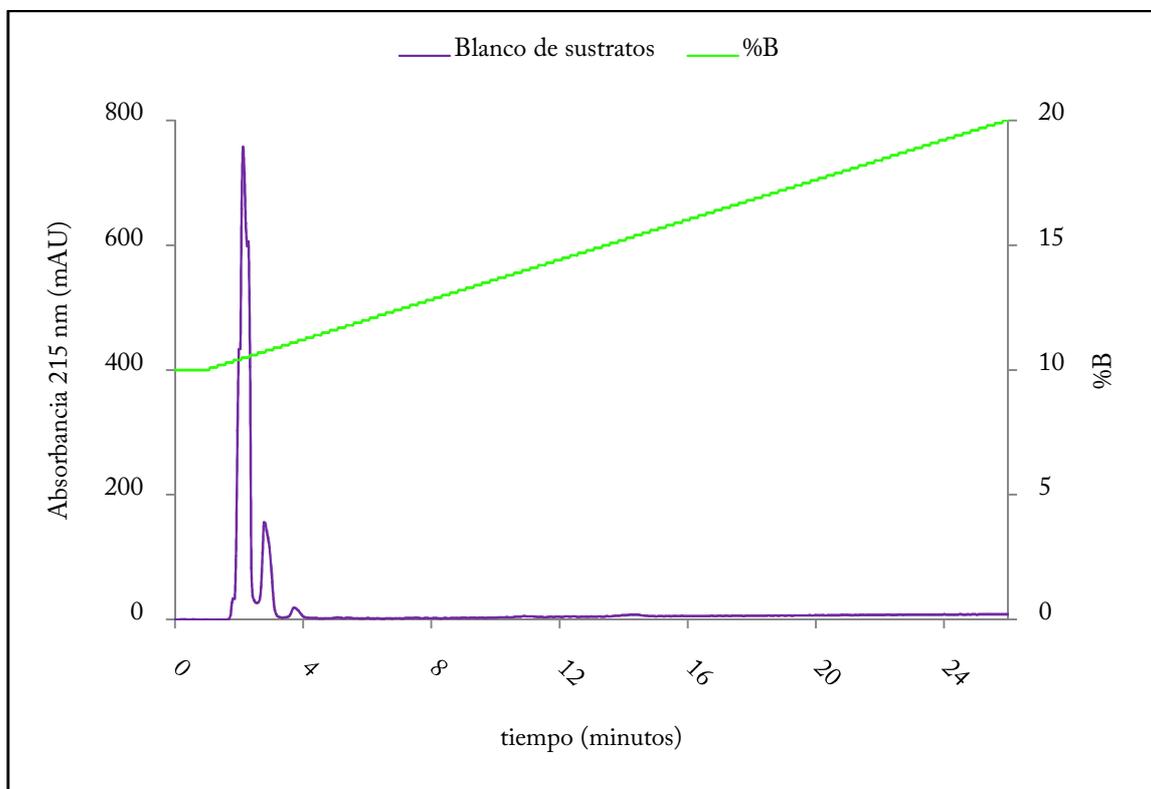


Figura 8.10: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en la mezcla ACN:H₂O (85:15). Cromatograma correspondiente al blanco de sustratos luego de 3 horas de reacción.

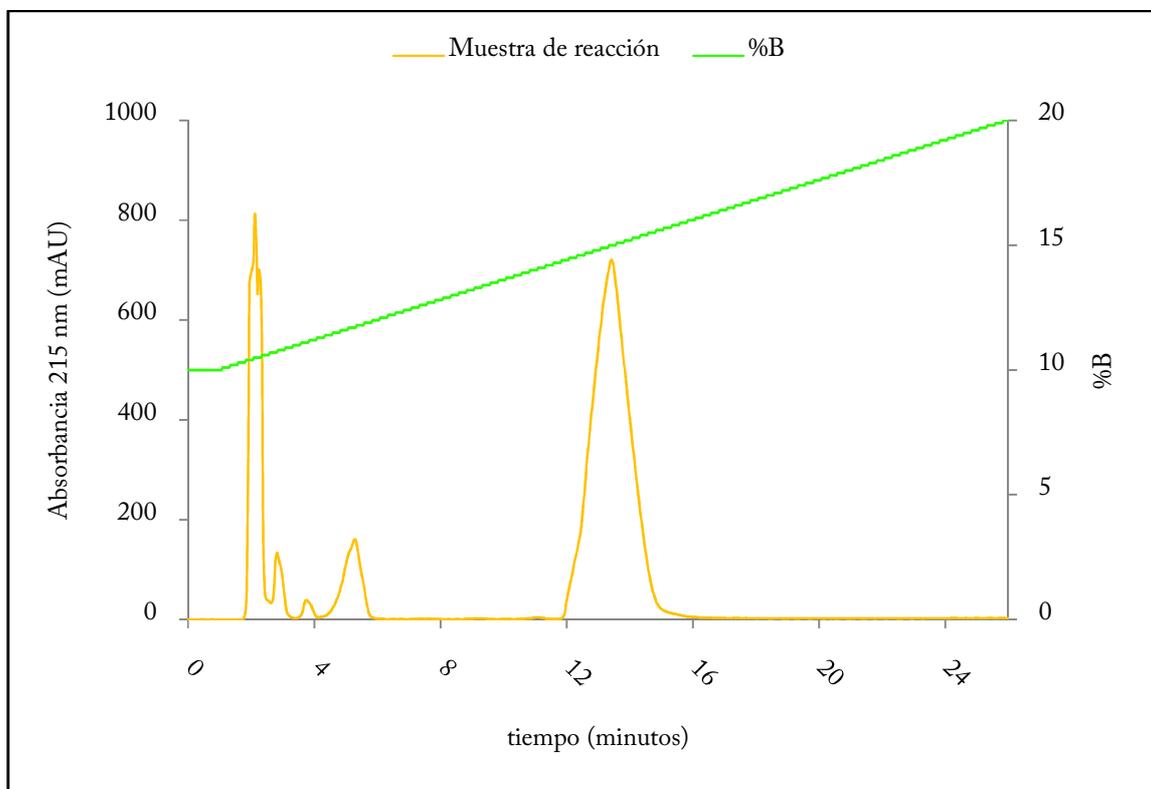


Figura 8.11: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en la mezcla ACN:H₂O (85:15). Cromatogramas correspondientes a la muestra de reacción, luego de 3 horas de reacción.

Medio EtOH:H₂O (85:15)

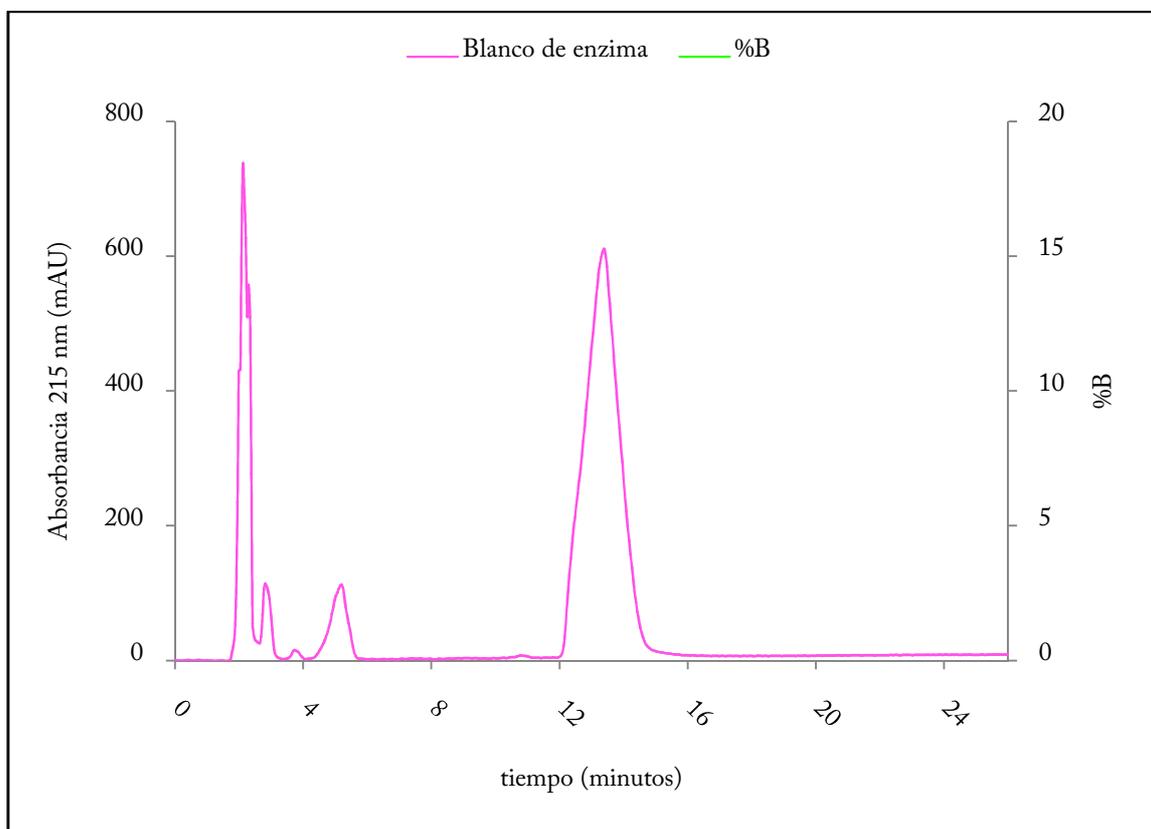


Figura 8.12: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en la mezcla EtOH:H₂O (85:15). Cromatogramas correspondiente al blanco de enzima luego de 3 horas de reacción.

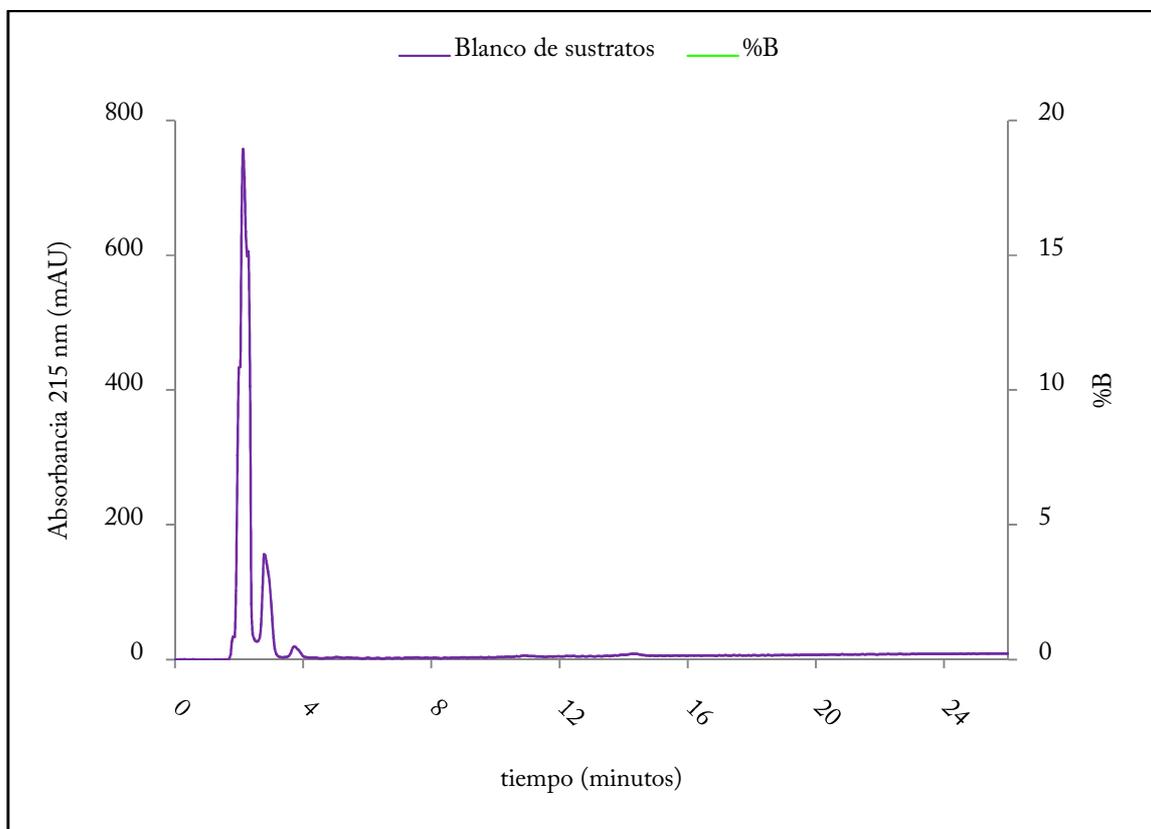


Figura 8.13: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en la mezcla EtOH:H₂O (85:15). Cromatograma correspondiente al blanco de sustratos luego de 3 horas de reacción.

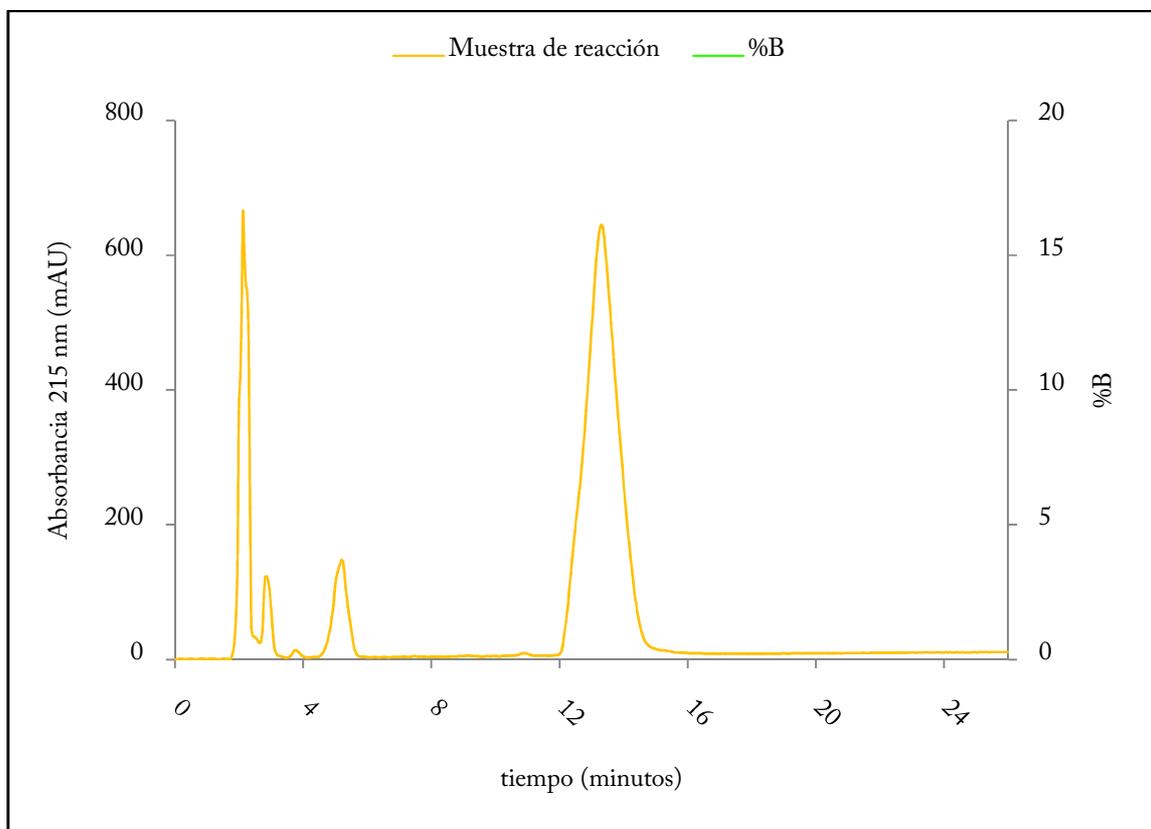


Figura 8.14: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicinol en la mezcla EtOH:H₂O (85:15). Cromatogramas correspondientes a la muestra de reacción, luego de 3 horas de reacción.

8.1.3. Síntesis enzimática de derivados de Bz-Arg-Gly utilizando glicina como nucleófilo.

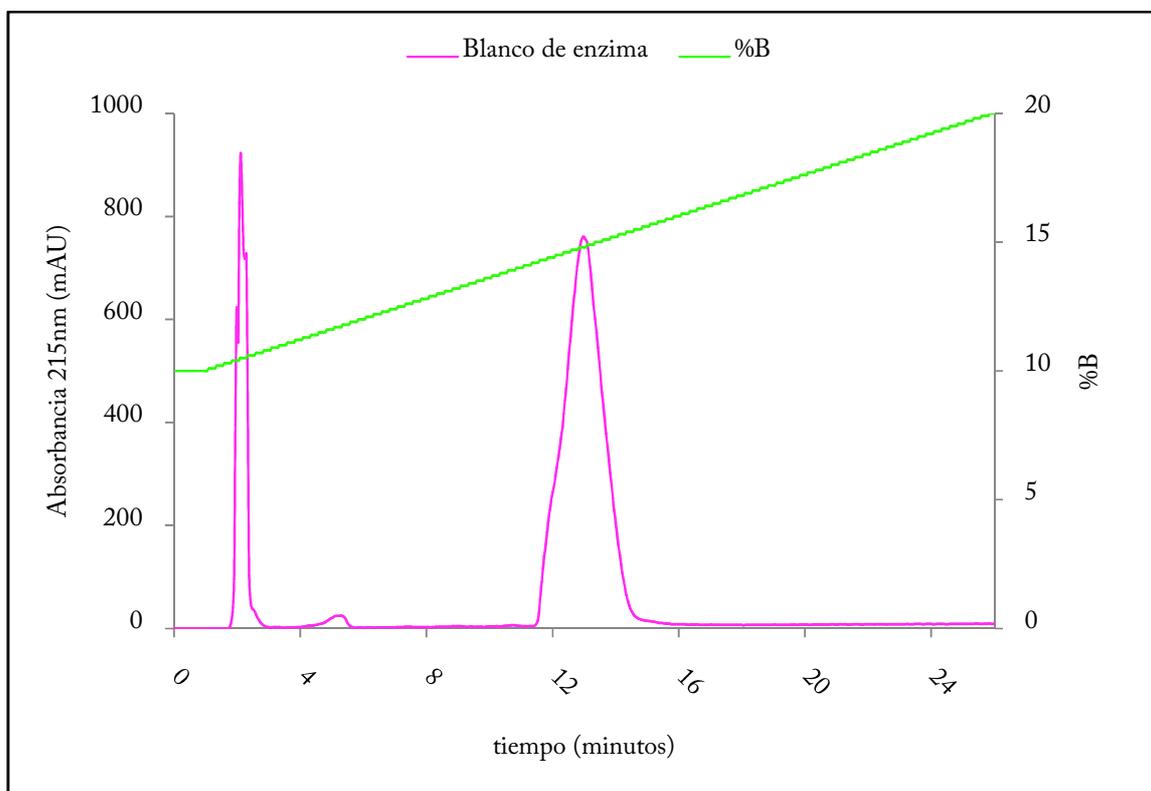


Figura 8.15: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicina en la mezcla ACN:H₂O (85:15). Cromatogramas correspondiente al blanco de enzima, luego de 3 horas de reacción.

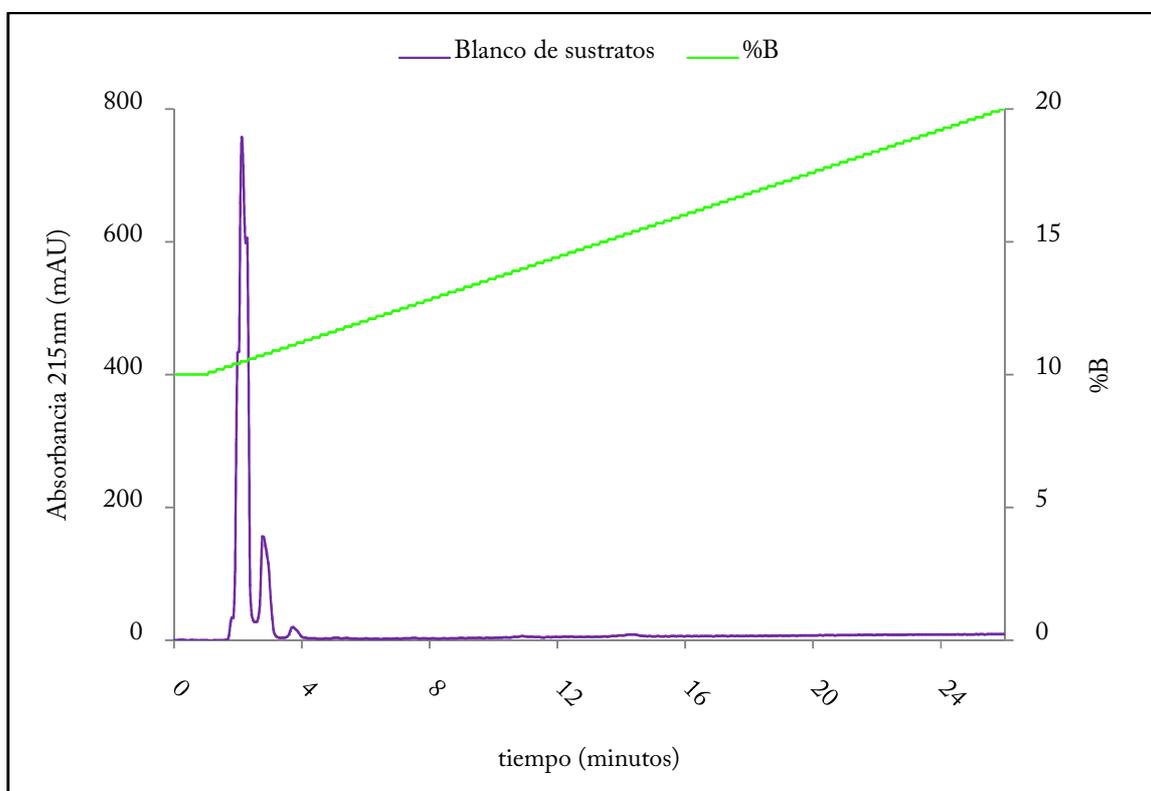


Figura 8.16: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicina en la mezcla ACN:H₂O (85:15). Cromatograma correspondiente al blanco de sustratos, luego de 3 horas de reacción.

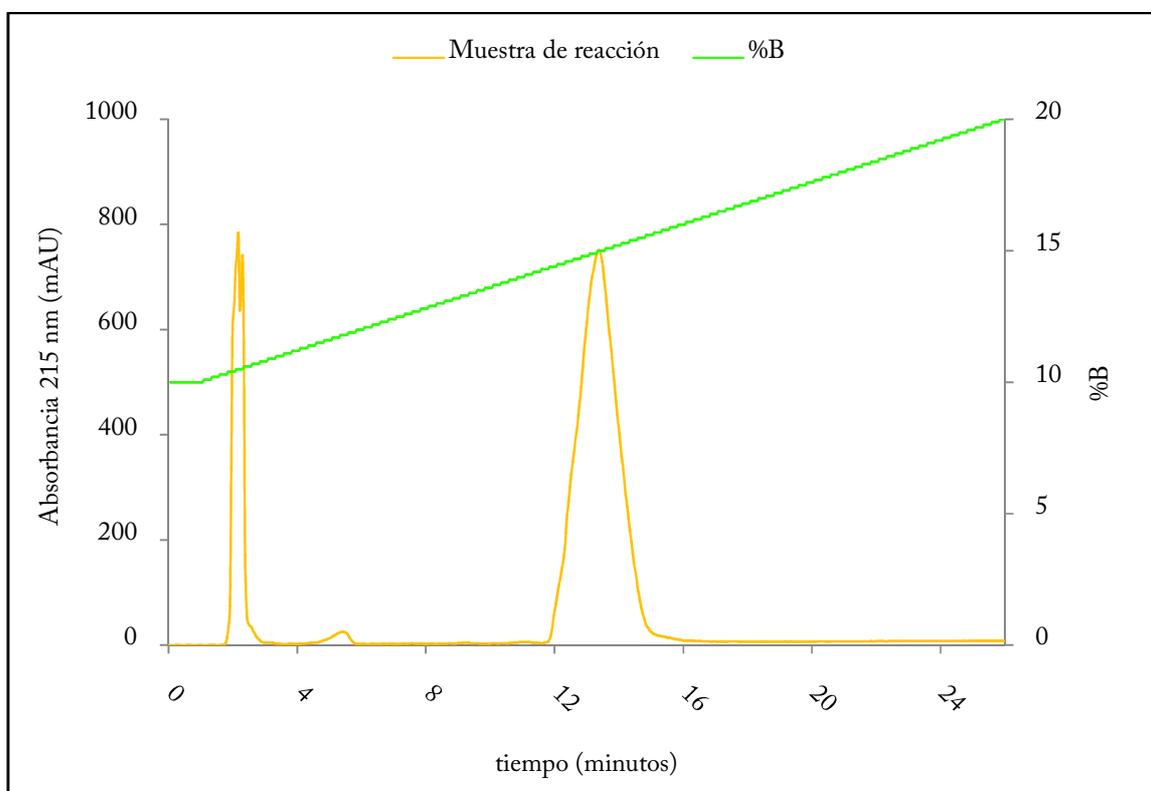


Figura 8.17: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con glicina en la mezcla ACN:H₂O (85:15). Cromatogramas correspondientes a la muestra de reacción, luego de 3 horas de reacción.

8.1.4. Síntesis enzimática de derivados de Bz-Arg-Ala-OMe utilizando Ala-OMe como nucleófilo.

Medio ACN:H₂O (85:15)

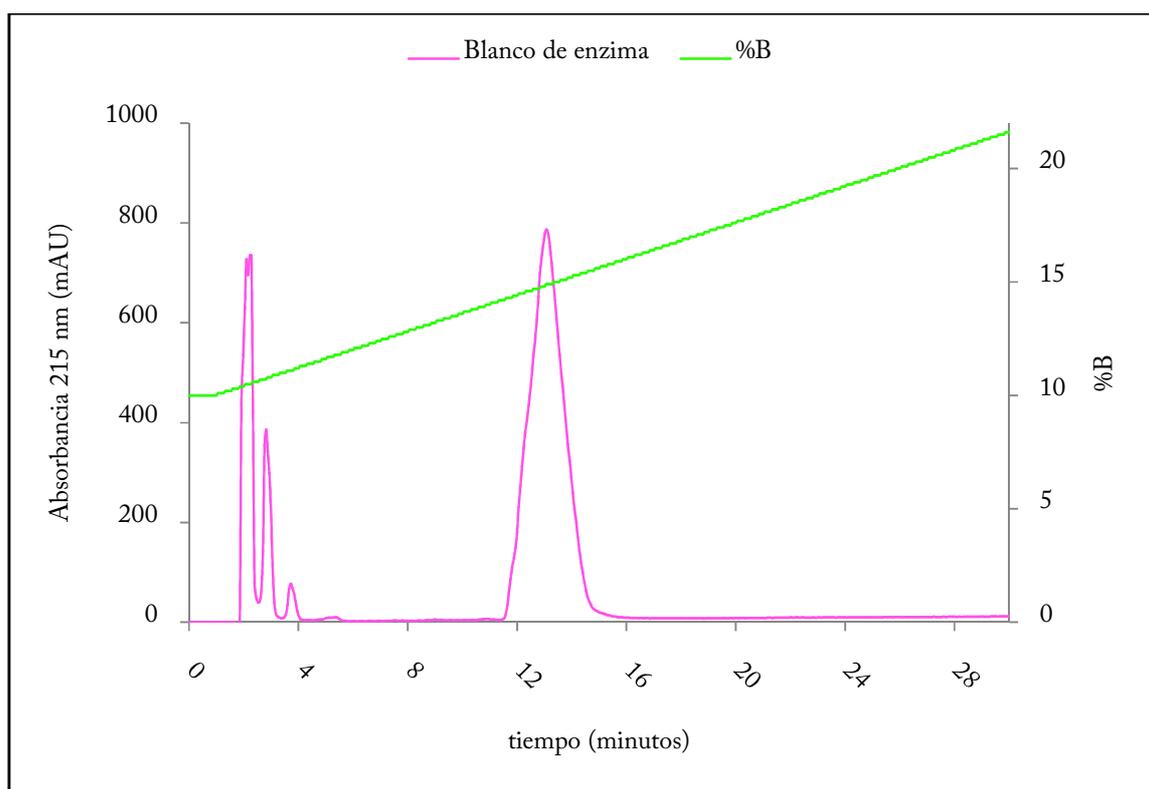


Figura 8.18: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla ACN:H₂O (85:15). Cromatograma correspondiente al blanco de enzima, luego de 3 horas de reacción.

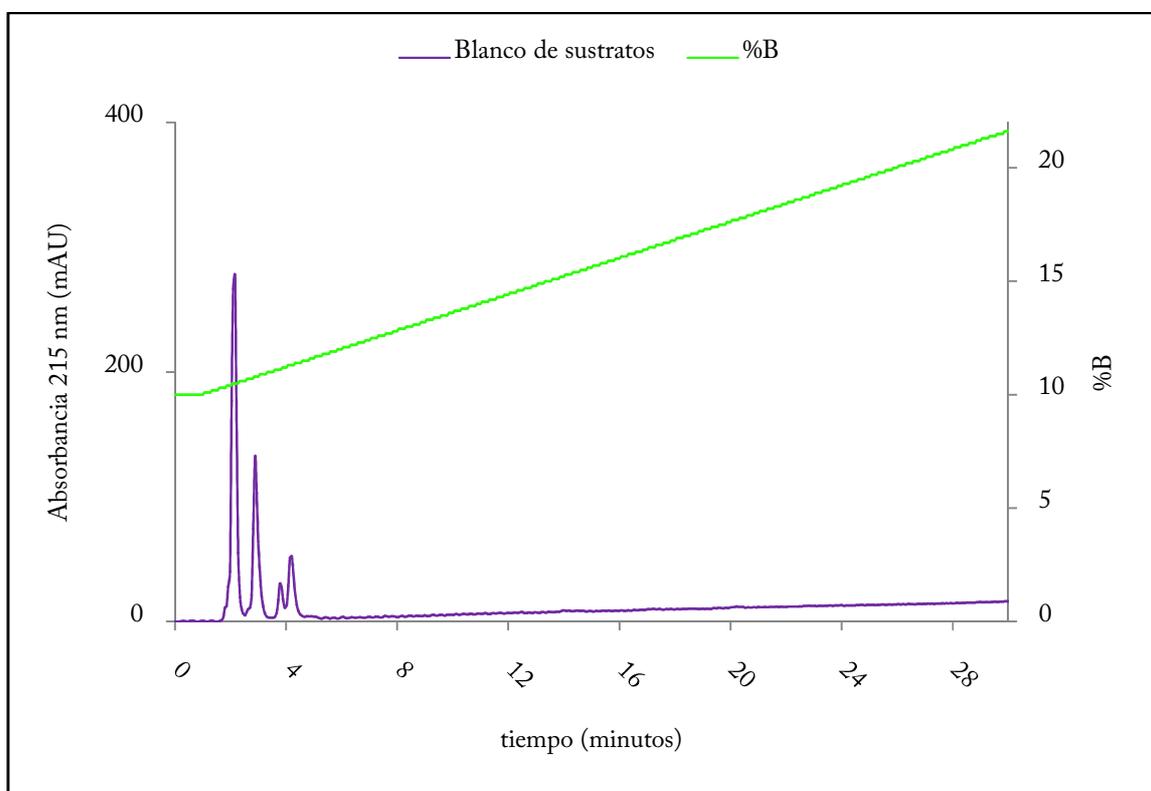


Figura 8.19: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla ACN:H₂O (85:15). Cromatograma correspondiente al blanco de sustratos, luego de 3 horas de reacción.

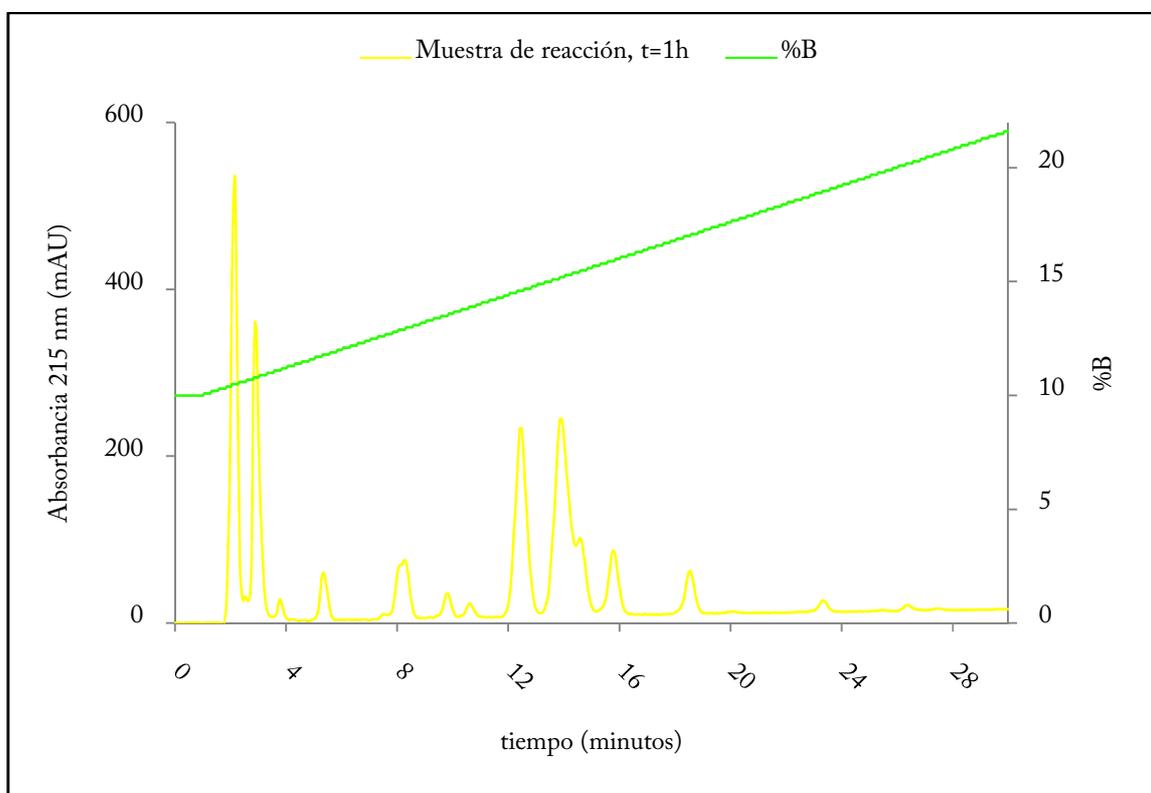


Figura 8.20: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla ACN:H₂O (85:15). Cromatograma correspondiente a la muestra de reacción, luego de 1 hora de reacción.

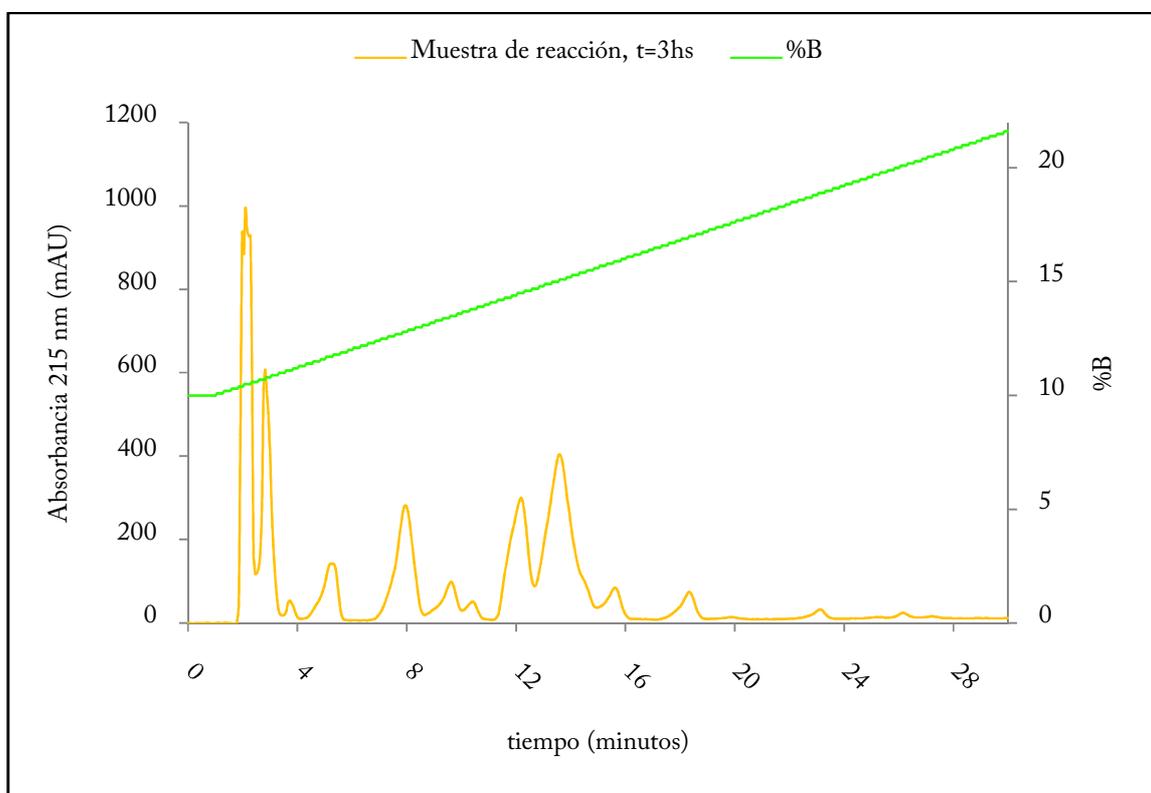


Figura 8.21: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla ACN:H₂O (85:15). Cromatograma correspondiente a la muestra de reacción, luego de 3 horas de reacción.

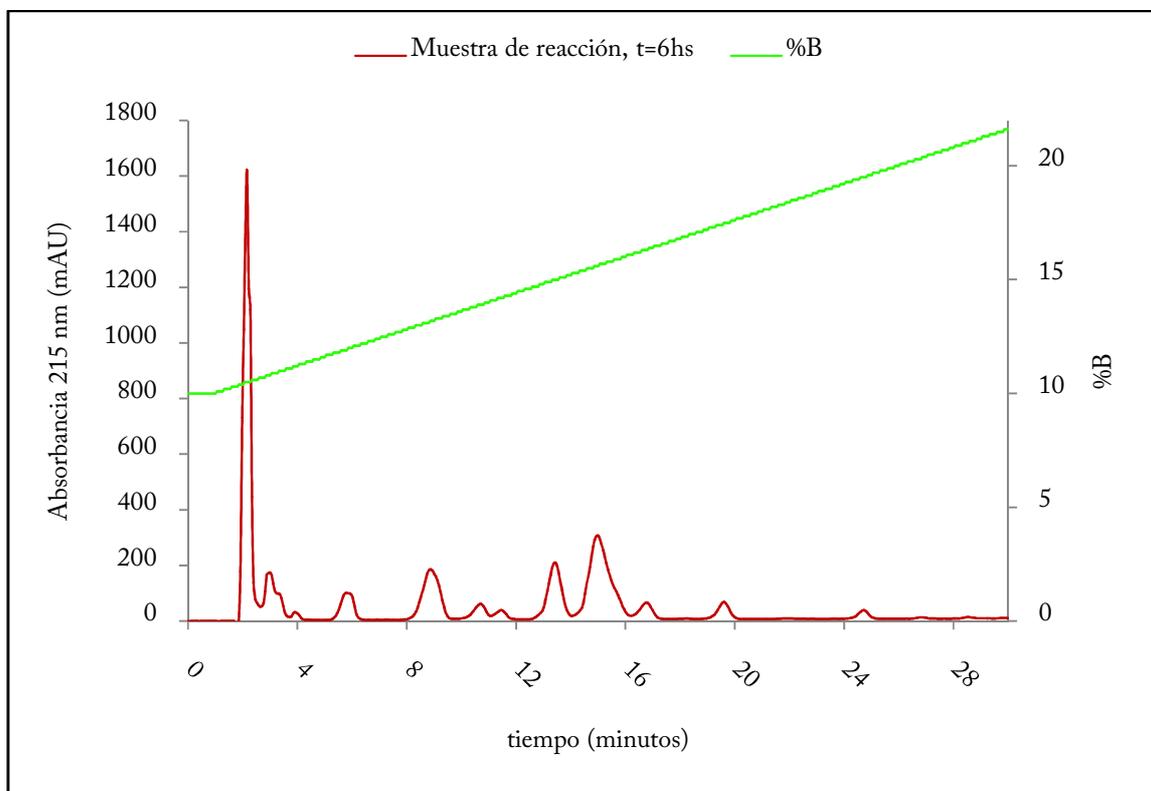


Figura 8.22: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla ACN:H₂O (85:15). Cromatograma correspondiente a la muestra de reacción, luego de 6 horas de reacción.

Medio EtOH:H₂O (85:15)

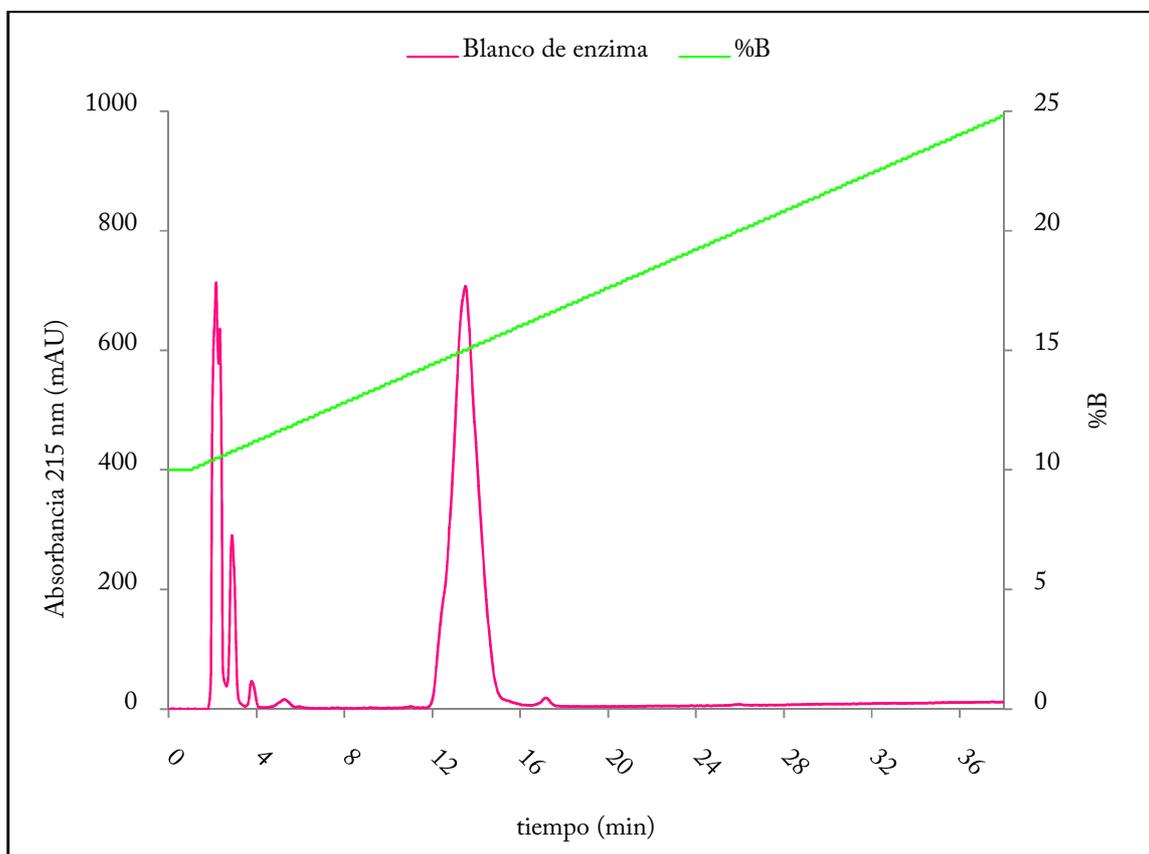


Figura 8.23: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla EtOH:H₂O (85:15). Cromatograma correspondiente al blanco de enzima, luego de 3 horas de reacción.

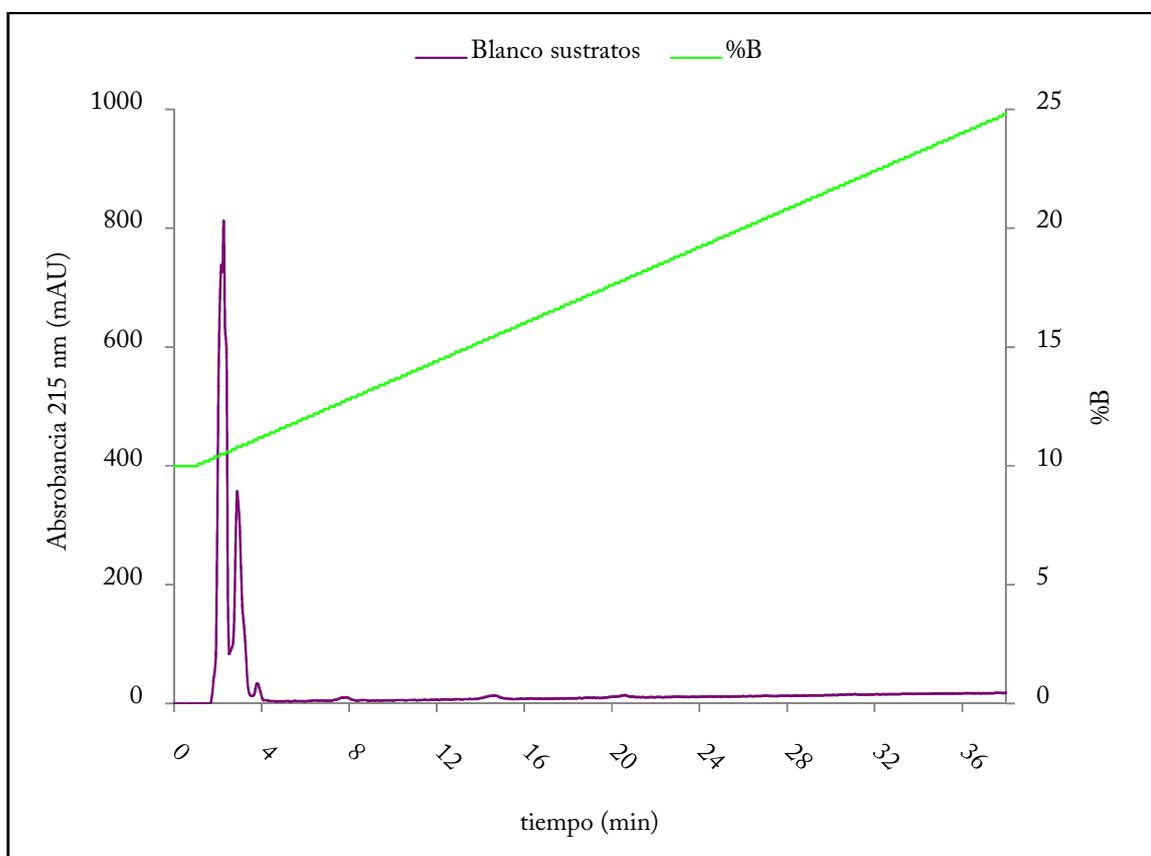


Figura 8.24: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla EtOH:H₂O (85:15). Cromatograma correspondiente al blanco de sustratos, luego de 3 horas de reacción.

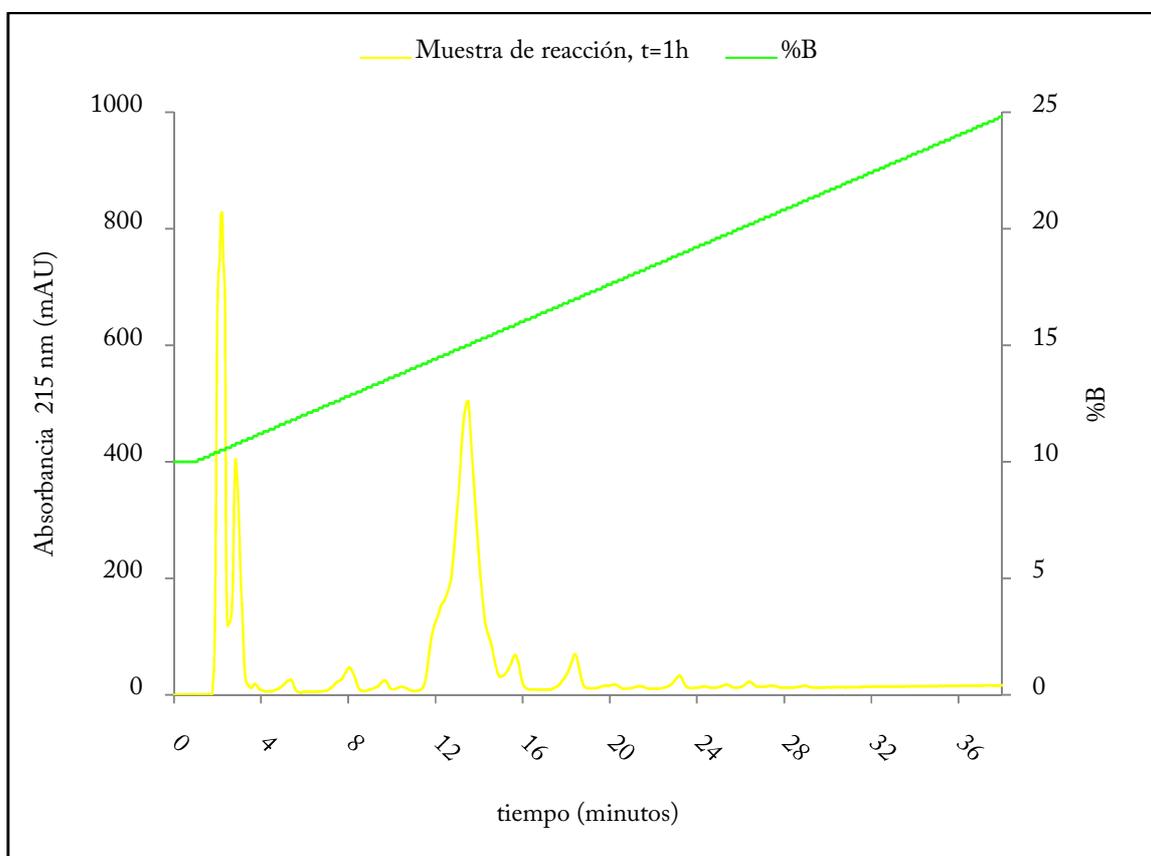


Figura 8.25: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla EtOH:H₂O (85:15). Cromatograma correspondiente a la muestra de reacción, luego de 1 hora de reacción.

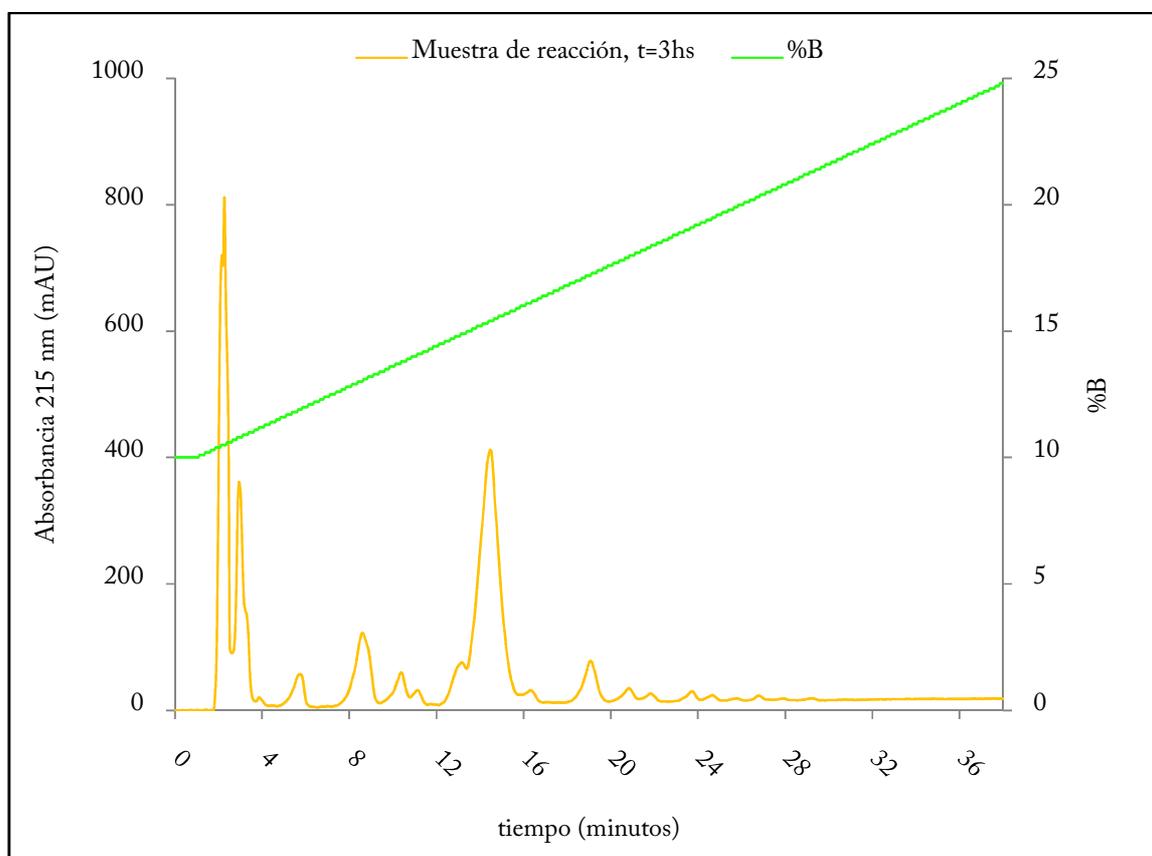


Figura 8.26: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla EtOH:H₂O (85:15). Cromatograma correspondiente a la muestra de reacción, luego de 3 horas de reacción.

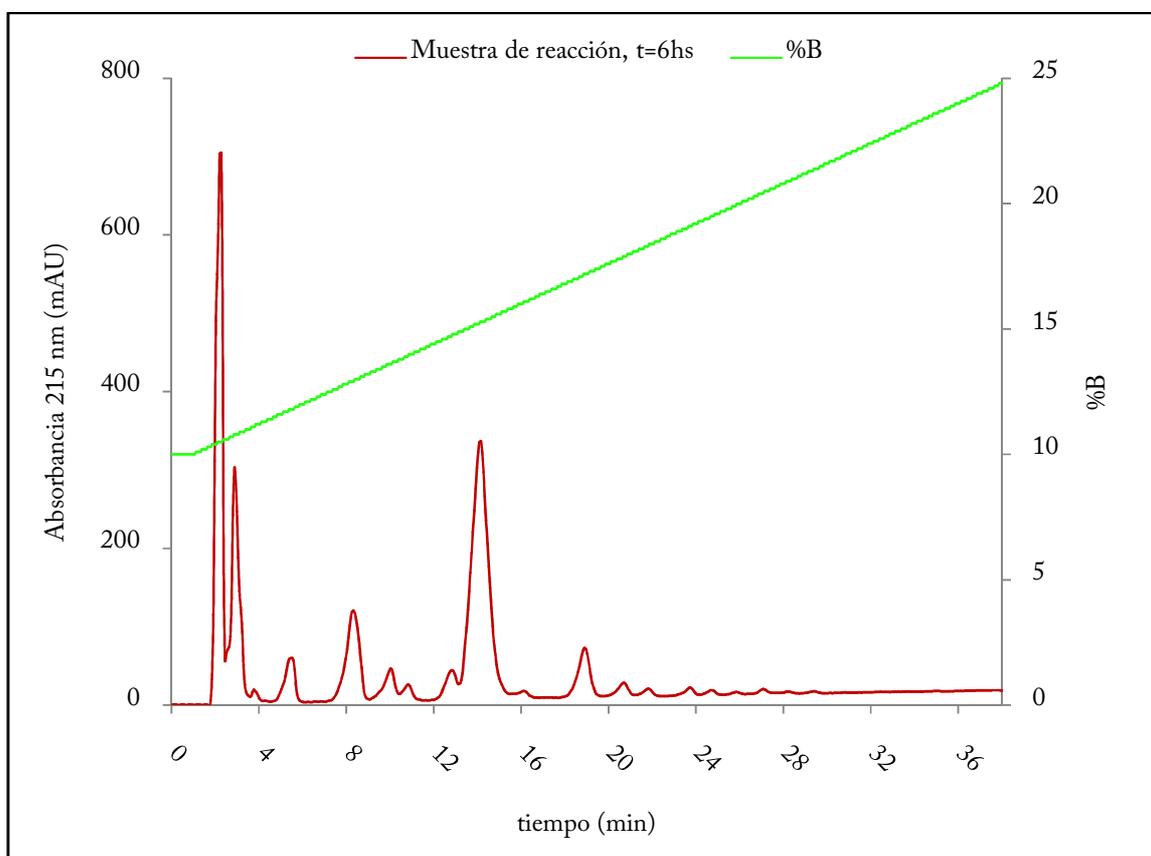


Figura 8.27: Análisis por RP-HPLC de la reacción Bz-Arg-OEt con Ala-OMe en la mezcla EtOH:H₂O (85:15). Cromatograma correspondiente a la muestra de reacción, luego de 6 horas de reacción.

8.1.5. Reportes del análisis por HPLC-ESI MS

Signal 1: DAD1 A, Sig=215,4 Ref=360,100

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	3.578	PV	0.2079	1.51468e4	982.49005	11.5141
2	3.875	VV	0.1383	5159.38330	526.60925	3.9220
3	4.083	VV	0.1582	5583.94629	516.08264	4.2448
4	4.307	VV	0.1835	1.06483e4	798.93488	8.0946
5	5.211	VV	0.3831	7486.49463	237.95357	5.6910
6	5.569	VV	0.2886	7300.39697	354.06134	5.5496
7	6.103	VB	0.2926	2973.09766	141.78094	2.2601
8	7.449	BP	0.2050	334.75848	24.25760	0.2545
9	10.921	BP	0.4459	4571.80273	147.39589	3.4754
10	16.354	BV	0.4747	5663.61572	162.46109	4.3053
11	16.936	VP	0.4332	4833.74170	141.97726	3.6745
12	20.177	BP	0.4095	1698.58289	53.62766	1.2912
13	25.583	BP	0.6291	8338.41211	158.66240	6.3386
14	28.633	BB	0.8862	1.32441e4	180.04367	10.0678
15	32.580	BP	0.4502	1360.10339	36.21254	1.0339
16	38.400	BP	0.4498	1093.42212	29.13908	0.8312
17	87.302	BV	0.9477	1.28947e4	169.46425	9.8022
18	87.547	VV	0.4916	6373.94141	157.67409	4.8453
19	88.403	VB	0.3045	2170.61987	95.72687	1.6500
20	88.884	BV	0.2061	1048.93140	68.69332	0.7974
21	89.251	VV	0.1965	2031.88354	137.39169	1.5446
22	89.480	VV	0.1763	1286.38965	98.53342	0.9779
23	89.831	VV	0.1771	2012.29919	161.81609	1.5297
24	90.349	VB	0.2643	1774.54980	90.67823	1.3490
25	90.896	BB	0.3041	701.52563	31.95946	0.5333
26	97.837	BB	0.3068	886.57196	41.24294	0.6739
27	99.952	BB	0.7169	4930.84326	81.60842	3.7483
Totals :				1.31549e5	5626.47864	

Figura 8.28: Reporte de áreas.

8.2. Aplicación de la metodología QSAR-3D para la obtención de modelos cualitativos de predicción de actividad sobre receptores IN de los derivados Bz-Arg

8.2.1. Conjunto de entrenamiento

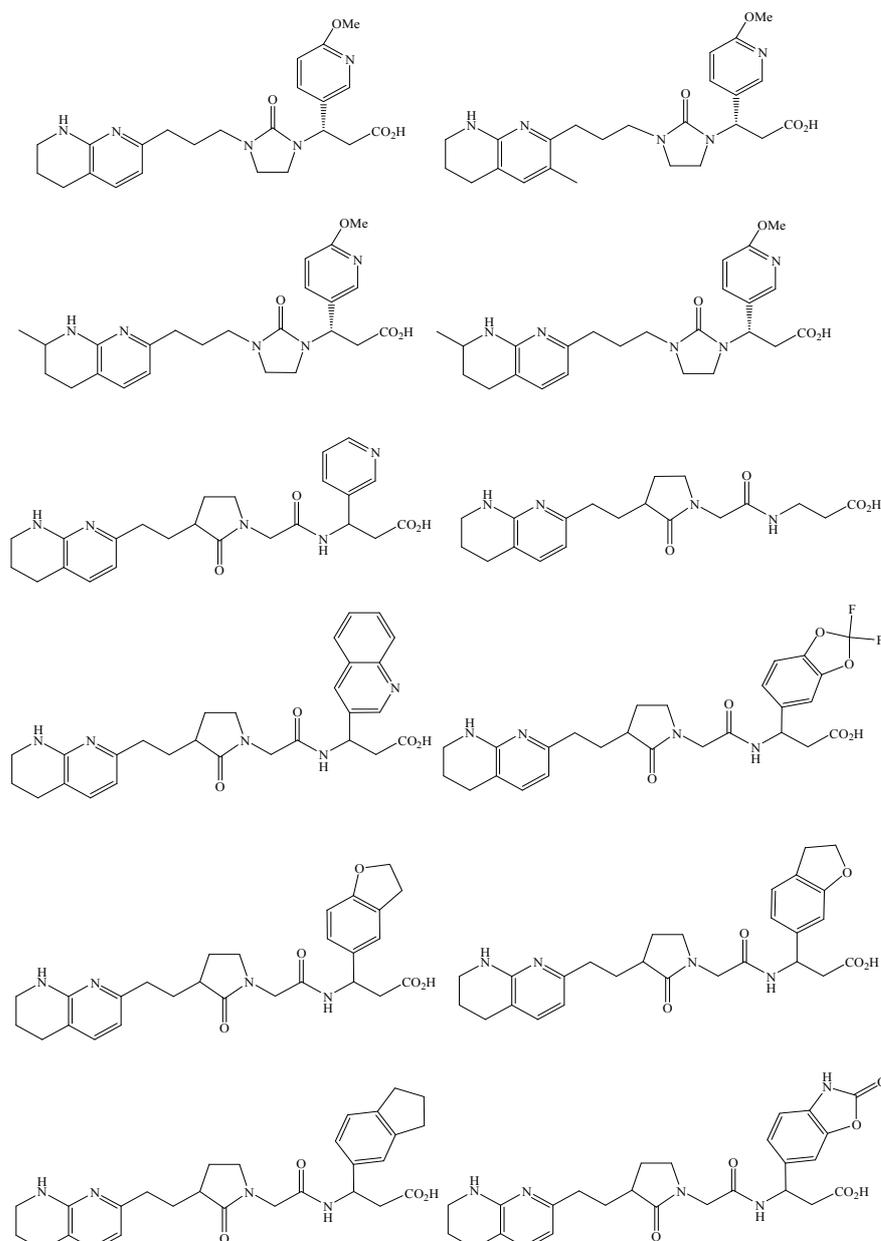


Figura 8.29: Compuestos activos utilizados en el conjunto de entrenamiento.

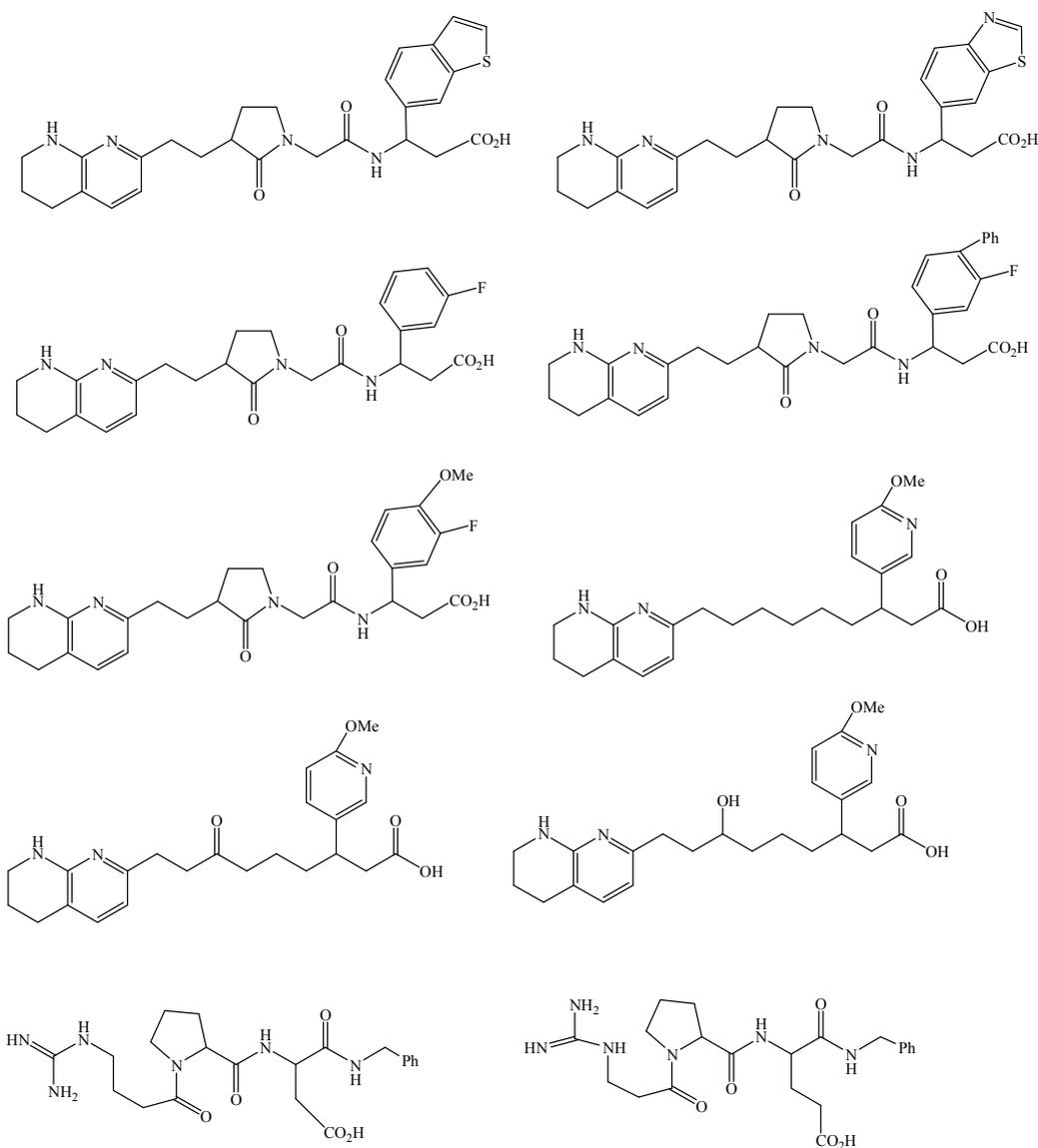


Figura 8.29 (continuación): Compuestos activos utilizados en el conjunto de entrenamiento.

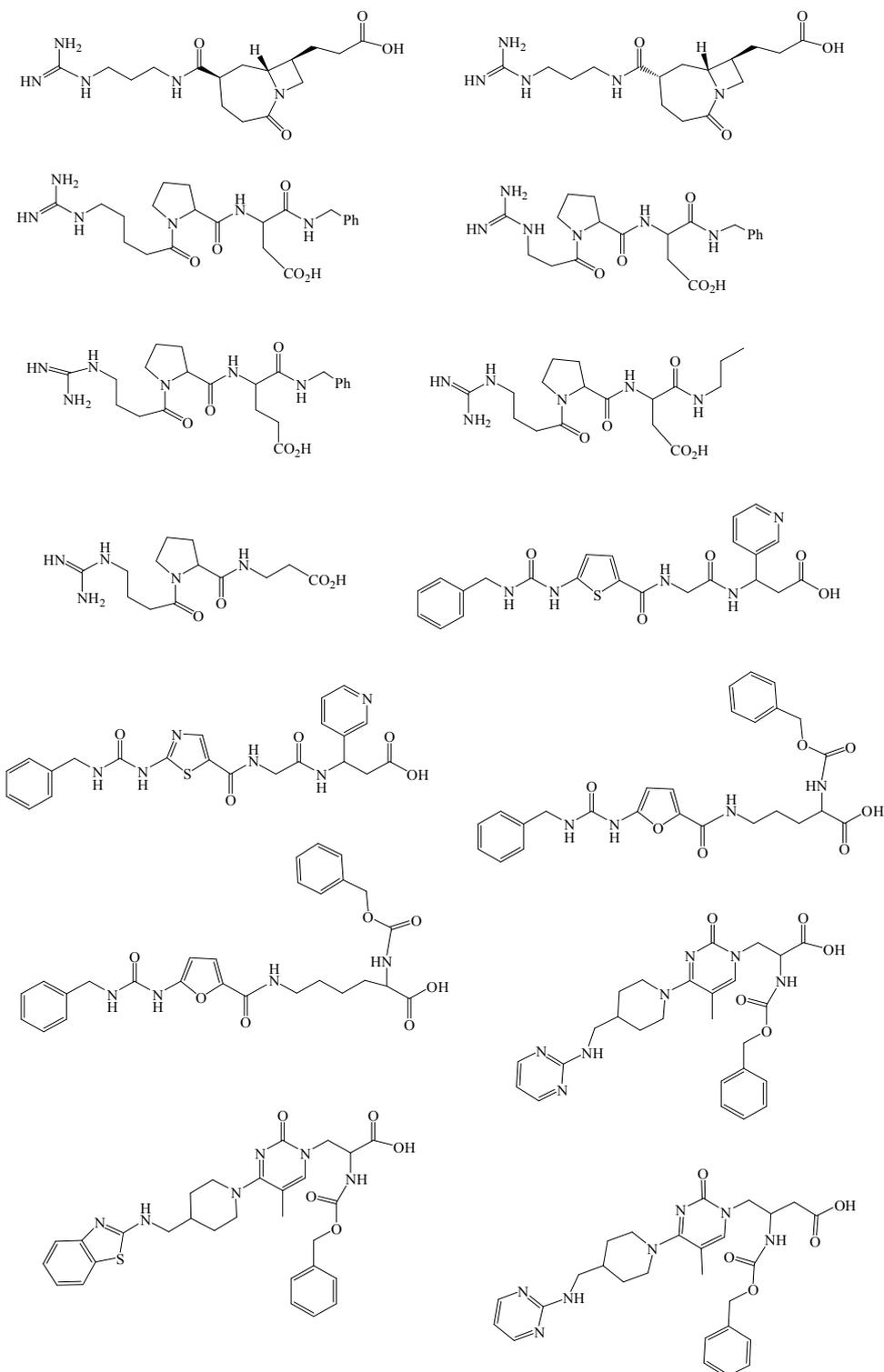


Figura 8.30: Compuestos inactivos utilizados en el conjunto de entrenamiento.

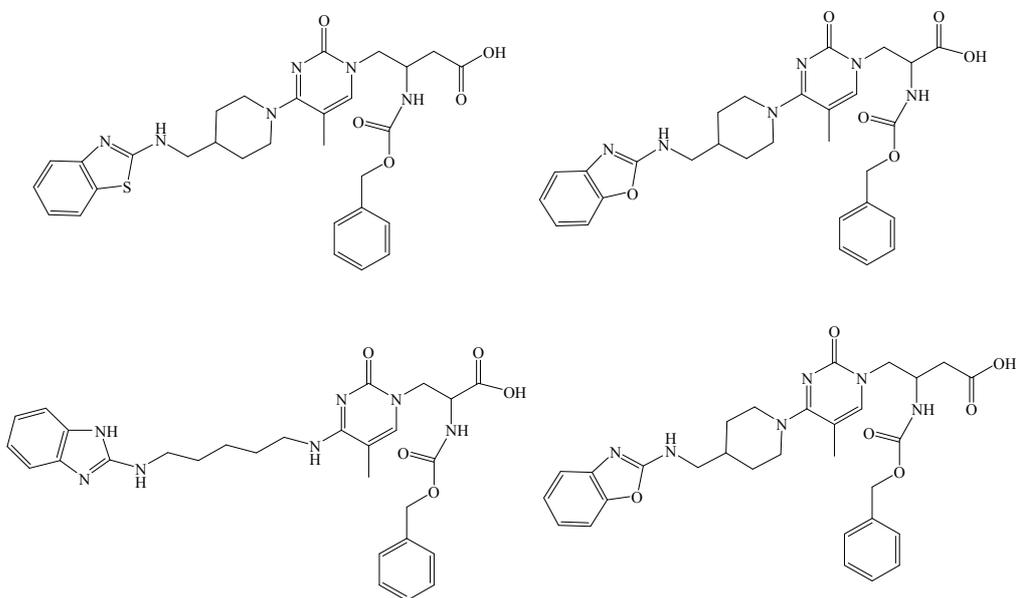


Figura 8.30 (continuación): Compuestos inactivos utilizados en el conjunto de entrenamiento.

8.2.2. Conjunto de prueba (validación externa)

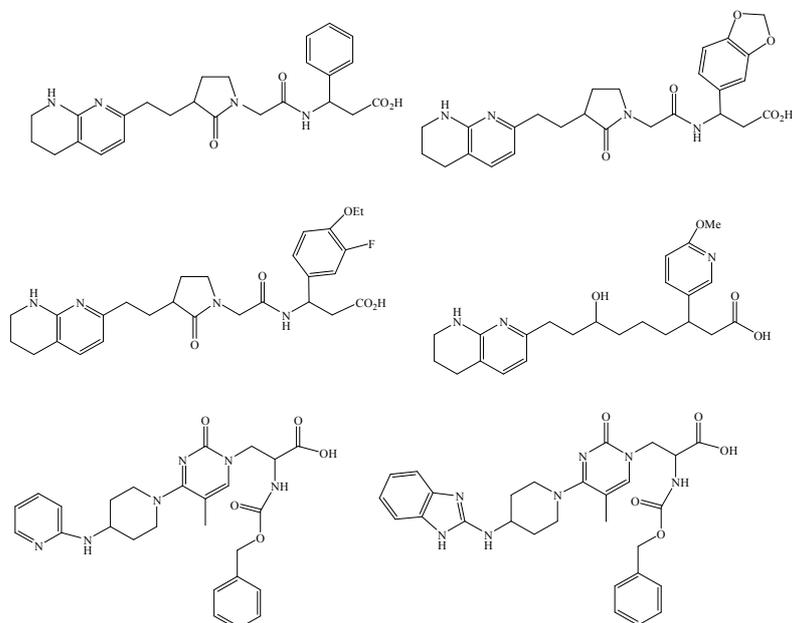


Figura 8.31: Compuestos activos utilizados en la validación externa.

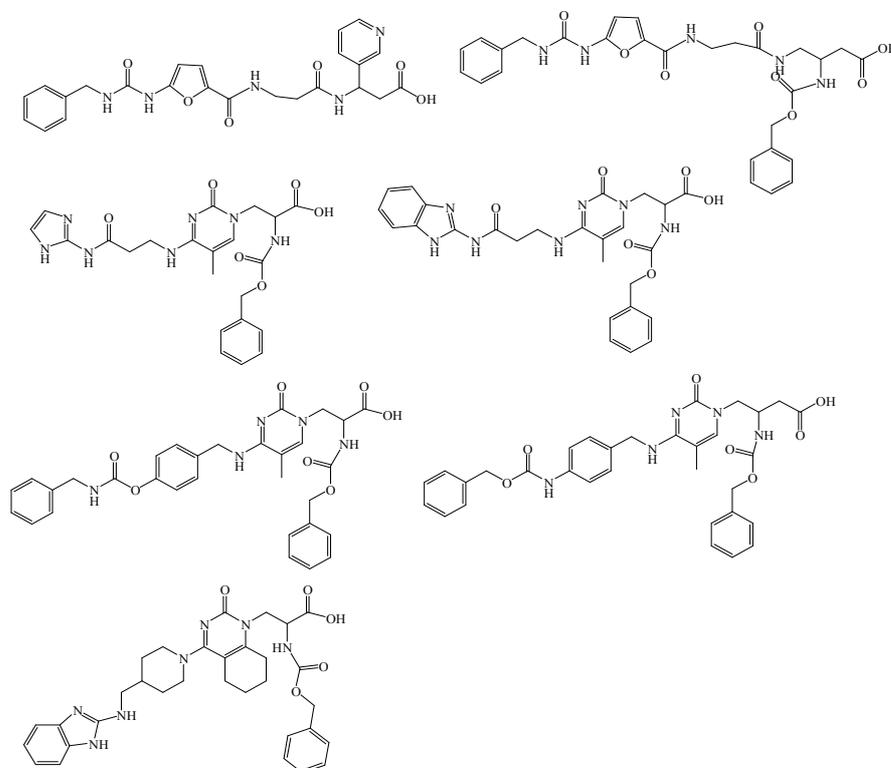


Figura 8.32: Compuestos inactivos utilizados en la validación externa.

8.2.3. Modelos Obtenidos

Modelo0314

	Beta	B	p-level
Constante		-0.08104	0.803426
Mor10p	0.551714	1.42942	0.000011
Mor11v	-0.350504	-1.08878	0.002477
Mor12m	0.246682	0.44752	0.020707

	% aciertos	INACTIVOS	ACTIVOS
INACTIVOS	88.88889	16	2
ACTIVOS	86.36364	3	19
Total	87.50000	19	21

F = 20.15061 p = .000000 Wilks' Lambda: .3732433

Modelo0513

	Beta	B	p-level
Constante		-3.55071	0.123340
RDF030p	0.414194	0.24114	0.000570
RDF020m	-0.450087	-3.61875	0.000181
IVDE	0.256944	2.97138	0.015764

	% aciertos	INACTIVOS	ACTIVOS
INACTIVOS	88.88889	16	2
ACTIVOS	90.90909	2	20
Total	90.00000	18	22

F = 21.69914 p = .000000 Wilks' Lambda: .3560922

Modelo0715

	Beta	B	p-level
Constante		-18.2420	0.000000
EEig02r	0.798708	5.5536	0.000000
As	-0.616373	-0.0230	0.000000
E3p	-0.197057	-3.3814	0.026403

	% aciertos	INACTIVOS	ACTIVOS
INACTIVOS	88.88889	16	2
ACTIVOS	95.45454	1	21
Total	92.50000	17	23

F = 34.66426 p = .000000 Wilks' Lambda: .2571561

Modelo030515

	Beta	B	p-level
Constante		25.820	0.000000
Uindex	-0.524901	-0.055	0.000121
Gu	-0.638572	-143.196	0.000006
Av	-0.420683	-0.014	0.004390

	% aciertos	INACTIVOS	ACTIVOS
INACTIVOS	88.88889	16	2
ACTIVOS	86.36364	3	19
Total	87.50000	19	21

F = 20.24125 p = .000000 Wilks' Lambda: .3721940

Modelo031015

	Beta	B	p-level
Constante		1.84134	0.005169
SEigZ	-0.594273	-1.52613	0.000010
piPC06	0.554843	0.00546	0.000028

	% aciertos	INACTIVOS	ACTIVOS
INACTIVOS	83.33334	15	3
ACTIVOS	86.36364	3	19
Total	85.00000	18	22

F = 20.33204 p = .000001 Wilks' Lambda: .4764108

Modelo1617

	Beta	B	p-level
Constante		-15.3158	0.000000
R2u	0.645432	6.8648	0.000000
nNHRPh	-0.447966	-1.3477	0.000019
nCt	0.311632	0.8287	0.001553

	% aciertos	INACTIVOS	ACTIVOS
INACTIVOS	88.88889	16	2
ACTIVOS	95.45454	1	21
Total	92.50000	17	23

F = 28.75016 p = .000000 Wilks' Lambda: .2944774

Modelo091320

Peptidasas vegetales como biocatalizadores para la obtención de intermediarios de síntesis de compuestos con potencial actividad farmacológica

Anexos

	Beta	B	p-level
Constante		1.95145	0.123069
RDF030p	0.349488	0.20347	0.005986
RDF020m	-0.440862	-3.54458	0.000271
RDF115m	-0.262595	-0.14194	0.025678

	% aciertos	INACTIVOS	ACTIVOS
INACTIVOS	88.88889	16	2
ACTIVOS	90.90909	2	20
Total	90.00000	18	22

F = 20.90089 p = .000000 Wilks' Lambda: .3647318