

UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

# **DESARROLLO DE UN MODELO COMPUTACIONAL PARA EL ESTUDIO DEL PROCESAMIENTO POSTRADUCCIONAL DE POTYVIRUS**

**Daniel Muñoz Escudero**

Universidad Nacional de Colombia

Facultad de Ciencias

Escuela de Biociencias

Medellín, Colombia

2017

# DESARROLLO DE UN MODELO COMPUTACIONAL PARA EL ESTUDIO DEL PROCESAMIENTO POSTRADUCCIONAL DE POTYVIRUS

DANIEL MUÑOZ ESCUDERO

Tesis presentada como requisito parcial para optar al título de:

**Magister en Ciencias**

Director:

PABLO GUTIÉRREZ SÁNCHEZ, Ph.D.

Codirector:

MAURICIO MARÍN MONTOYA, Ph.D.

Línea de Investigación:

BIOTECNOLOGÍA

Universidad Nacional de Colombia

Facultad de Ciencias

Escuela de Biociencias

Medellín, Colombia

2017

# Agradecimientos

Agradezco a la Universidad Nacional de Colombia por la financiación brindada para la ejecución de este proyecto.

También a mi familia por motivarme a seguir una vida dedicada a la educación y por infundir la disciplina y el apoyo para afrontarla.

A mis amigos dentro y fuera del 19A-312, de quienes aprecio profundamente su ayuda y participación en esta tesis.

Finalmente, doy gracias a mis directores Pablo Gutiérrez y Mauricio Marín por su acompañamiento y total entrega en su labor docente.

# Resumen

El género *Potyvirus* incluye un amplio grupo de virus de plantas que afectan diversos cultivos de gramíneas y dicotiledóneas en Colombia y el mundo. Su especie tipo, el *Potato virus Y* (PVY), es uno de los principales virus de distribución global que afecta severamente diferentes plantas solanáceas de importancia económica como la papa y el tomate, lo que ha promovido el desarrollo de investigaciones en métodos de detección, manejo y comprensión de su ciclo infectivo. La secuenciación de nueva generación (NGS) ha permitido establecer estrategias para la identificación rápida de especies virales, aún en condiciones de bajo título viral, y que en conjunto con aplicaciones bioinformáticas, representan una herramienta fundamental para el apoyo de los estudios de detección asintomática, relaciones filogenéticas y biología de los virus. Sobre este último aspecto, se destacan los estudios dirigidos a la comprensión de los mecanismos de replicación y expresión de los genomas virales, específicamente en las etapas que involucran interacciones proteína-proteína, tales como el procesamiento postraduccional de las poliproteínas potyvirales.

En este trabajo se presenta la caracterización molecular de aislamientos de PVY infectando plantas de *Solanum tuberosum* subsp. andigena var. Diacol-Capiro en las dos principales regiones cultivadoras de papa de Antioquia (Oriente y Norte), utilizando metodologías de biología molecular como RT-PCR, pruebas de ELISA, secuenciación Sanger y NGS. Dicha información fue posteriormente empleada para proponer un modelo cinético de maduración de la poliproteína de PVY, a partir de simulaciones por Dinámica Molecular y tomando como referencia la estructura cristalográfica de la proteasa NIa-Pro del potyvirus *Tobacco etch virus* (TEV). La ejecución del trabajo permitió obtener un panorama de los niveles de diversidad genómica y de incidencia de PVY en las principales regiones cultivadoras de papa en Antioquia. Adicionalmente, la aplicación del modelo desarrollado aporta nueva información básica sobre el procesamiento postraduccional de este virus, un aspecto fundamental de la biología de los potyvirus. Se espera que estos resultados en su conjunto, apoyen los programas de manejo integrado de las enfermedades causadas por PVY en papa y otros cultivos en el país.

**Palabras clave:** Dinámica Molecular, ELISA, Potyvirus, Secuenciación de Nueva Generación

## Abstract

The genus *Potyvirus* comprises a large number of plant viruses that affect several gramineous and dicotyledonous crops worldwide. *Potato virus Y* (PVY), the type species of the genus, is one the most important viruses affecting solanaceous crops of economic importance around the world like potato and tomato. With the advent of Next-generation sequencing (NGS) technologies, it is now possible to quickly identify the presence of viruses infecting plants, even at low titers. In combination with bioinformatic analysis tools, NGS has become an important method for studying the basic biology of viruses, their evolutionary relationships and as a detection tool in asymptomatic plants.

This work presents the molecular characterization of two PVY isolates infecting *Solanum tuberosum* subsp. *andigena* var. *Diacol-Capiro* in the two main potato-producing regions of Antioquia using RT-PCR, ELISA, Sanger sequencing and NGS. The genomic data was used to model the dynamic behaviour of the NIa-Pro protease in presence of each of its cognate cleavage sites and predict the maturation kinetics of each PVY isolate. This model was first tested using the X-ray structure of the NIa-Pro protein *Tobacco etch virus* (TEV) and validated using published experimental data. This investigation gives an overview on the genomic diversity and incidence levels of PVY in potato-producing regions of Antioquia. Upon experimental validation, the theoretical methods developed in this work could be used after NGS sequencing to make predictions on the infection cycle of potyviruses. The results presented here will be useful for integrated disease management programs against PVY in potato and other solanaceous crops.

**Key words:** ELISA, Molecular Dynamics, Next-Generation Sequencing, Potyvirus

# Contenido

<b>RESUMEN</b>	<b>4</b>
<b>ABSTRACT</b>	<b>5</b>
<b>INTRODUCCIÓN</b>	<b>10</b>
<b>1. MARCO TEÓRICO</b>	<b>14</b>
<b>1.1. <i>Potato virus Y</i> (PVY)</b>	<b>14</b>
1.1.1. Diversidad del <i>Potato virus Y</i>	17
<b>1.2. Enfermedades del cultivo de la papa</b>	<b>20</b>
1.2.1. Cultivo de papa en Colombia: Incidencia de PVY	25
<b>1.3. Métodos de manejo de enfermedades virales</b>	<b>27</b>
<b>1.4. Biología molecular de PVY</b>	<b>29</b>
1.4.1. Procesamiento de la poliproteína en PVY	32
<b>1.5. Métodos de Dinámica Molecular en los estudios de virus</b>	<b>34</b>
<b>2. BIBLIOGRAFÍA</b>	<b>39</b>
<b>3. OBJETIVOS</b>	<b>55</b>
<b>3.1. General</b>	<b>55</b>
<b>3.2. Específicos</b>	<b>55</b>
<b>4. RESULTADOS</b>	<b>56</b>
<b>4.1. Capítulo 1. Análisis del mecanismo de reconocimiento de N1a-Pro para la postulación de un modelo del procesamiento postraduccional de la poliproteína en TEV y PVY, utilizando simulaciones por Dinámica Molecular.</b>	<b>56</b>
4.1.1. Introducción	56
4.1.2. Descripción del mecanismo de corte: paralelo entre N1a-Pro y tripsina (serín-proteasa)	56
4.1.3. Modelación por homología	62
4.1.4. Simulaciones por Dinámica Molecular (MD)	64
4.1.5. Resultados de las simulaciones por Dinámica Molecular	72
4.1.6. Modelo cinético	77
4.1.7. Análisis de las frecuencias de corte en la poliproteína de TEV	84
4.1.8. Modelo de maduración cinética en el sistema PVY	98

4.1.9. Bibliografía	115
4.1.10. ANEXOS – Capítulo 1	118
<b>4.2. Capítulo 2. Detección y caracterización molecular del <i>Potato virus Y</i> (PVY) en cultivos de papa del Norte de Antioquia (Colombia)</b>	<b>127</b>
4.2.1. ANEXOS – Capítulo 2	154
<b>4.3. Capítulo 3. Genome characterization of <i>Potato virus Y</i> strains infecting potato (<i>Solanum tuberosum</i> L.) in La Unión (Antioquia, Colombia)</b>	<b>160</b>
4.3.1. ANEXOS – Capítulo 3	182
<b>5. ANEXOS GENERALES</b>	<b>187</b>
5.1. Análisis bioinformático para el ensamble de virus mediante datos de secuenciación por NGS	187
5.2. Script <i>mapeador.pl</i>	188
5.3. Identificación de polimorfismos	191
<b>6. CONCLUSIONES</b>	<b>192</b>

# Lista de figuras

	Pág.
<b>Figura 1.</b> Características del genoma de PVY.	15
<b>Figura 2.</b> Principales síntomas inducidos por PVY.	17
<b>Figura 3.</b> Filogenia simplificada de PVY.	18
<b>Figura 4.</b> Patrones de recombinación entre dos razas de PVY.	20
<b>Figura 5.</b> Microorganismos asociados a las principales enfermedades en papa no causadas por virus.	21
<b>Figura 6.</b> Síntomas inducidos por diversas especies virales en papa.	22
<b>Figura 7.</b> Diagrama de generaciones en la producción de semilla de papa.	28
<b>Figura 8.</b> Características relevantes de la poliproteína potyviral.	32
<b>Figura 9.</b> Modelo de maduración enzimática para un virus de tres componentes.	37
<b>Figura 10.</b> Efecto de las constantes de afinidad sobre la cinética de maduración de un virus hipotético con una poliproteína de tres componentes P/R/C.	38
<b>Figura 11.</b> Proteasa NIa-Pro catalíticamente activa de <i>Tobacco etch virus</i> .	57
<b>Figura 12.</b> Residuos de la tríada catalítica de NIa-Pro de TEV.	58
<b>Figura 13.</b> Residuos de la tríada catalítica de las serín y cisteín-proteasas.	59
<b>Figura 14.</b> Mecanismo catalítico postulado para la NIa-Pro de TEV.	60
<b>Figura 15.</b> Estabilización del estado de transición en la NIa-Pro de TEV.	61
<b>Figura 16.</b> Geometría del carbono carbonílico en el enlace peptídico.	62
<b>Figura 17.</b> Archivos requeridos por MODELLER para la modelación por homología.	62
<b>Figura 18.</b> Secuencias de las regiones de corte para TEV y PVY.	63
<b>Figura 19.</b> Archivo de reporte de MODELLER.	63
<b>Figura 20.</b> Formato de los archivos PDB y PSF utilizados en las simulaciones moleculares.	64
<b>Figura 21.</b> Formato del archivo de topología para la definición de la prolina.	65
<b>Figura 22.</b> Posibles estados de protonación de la histidina.	66
<b>Figura 23.</b> Definición de los estados de protonación de la histidina.	67
<b>Figura 24.</b> Caja de solvatación.	68
<b>Figura 25.</b> Diagrama de flujo para las simulaciones por MD en NAMD.	69
<b>Figura 26.</b> Principales secciones del archivo de configuración.	70
<b>Figura 27.</b> Proceso de la simulación por Dinámica Molecular.	71
<b>Figura 28.</b> Patrón de distribución de los átomos de azufre de la Cys-144 en diversos momentos de la simulación.	73
<b>Figura 29.</b> Distribución de los sitios de corte simulados.	74
<b>Figura 30.</b> Configuración A, tomando como referencia el corte NIb/CP.	75
<b>Figura 31.</b> Configuración B.	76
<b>Figura 32.</b> Configuración C.	77
<b>Figura 33.</b> Nomenclatura usada en el modelo enzimático.	78
<b>Figura 34.</b> Matriz de especies $E(t)$ .	81
<b>Figura 35.</b> Cinética del dipéptido P <sub>45</sub> (6K2-VPg).	82
<b>Figura 36.</b> Cinética de maduración en escenarios inter e intramoleculares.	83
<b>Figura 37.</b> Simulación de Western blot para especies con CP.	84
<b>Figura 38.</b> Frontera de decisión según el corte NIb/CP de TEV.	85
<b>Figura 39.</b> Sitio de corte NIb/CP.	86
<b>Figura 40.</b> Sitio de corte P1/HC-Pro.	87
<b>Figura 41.</b> Sitio de corte HC-Pro/P3.	88

<b>Figura 42.</b> Sitio de corte P3/6K1.	89
<b>Figura 43.</b> Sitio de corte 6K1/CI.	90
<b>Figura 44.</b> Sitio de corte CI/6K2.	91
<b>Figura 45.</b> Sitio de corte 6K2/VPg.	92
<b>Figura 46.</b> Sitio de corte VPg/Nla-Pro.	93
<b>Figura 47.</b> Sitio de corte Nla-Pro/Nlb.	94
<b>Figura 48.</b> Simulación de las mutaciones en el sitio de corte Nlb/CP.	97
<b>Figura 49.</b> Tiempo del procesamiento de la poliproteína Nlb-CP.	98
<b>Figura 50.</b> Análisis de secuencias para los sitios de corte de f_PVY y t_PVY.	99
<b>Figura 51.</b> Sitio de corte P1/HC-Pro del sistema PVY.	100
<b>Figura 52.</b> Sitio de corte HC-Pro/P3 del sistema PVY.	101
<b>Figura 53.</b> Sitio de corte P3/6K1 del sistema PVY.	102
<b>Figura 54.</b> Sitio de corte 6K1/CI del sistema PVY.	103
<b>Figura 55.</b> Sitio de corte CI/6K2 del sistema PVY.	104
<b>Figura 56.</b> Sitio de corte 6K2/VPg del sistema PVY.	104
<b>Figura 57.</b> Sitio de corte VPg/Nla-Pro del sistema PVY.	105
<b>Figura 58.</b> Sitio de corte Nla-Pro/Nlb del sistema PVY.	106
<b>Figura 59.</b> Paralelo entre las regiones de corte Nlb/CP en el sistema PVY.	107
<b>Figura 60.</b> Alineamiento de secuencias de la proteasa Nla-Pro en f_PVY y t_PVY.	109
<b>Figura 61.</b> Análisis del movimiento local de las proteasas Nla-Pro en el sistema PVY.	111
<b>Figura 62.</b> Resultados de la cinética para el sistema PVY.	113

## Lista de tablas

	<b>Pág.</b>
<b>Tabla 1.</b> Respuestas sintomáticas y resistencia hipersensible según las variantes de PVY.	19
<b>Tabla 2.</b> Principales virus reportados en cultivos de papa a nivel mundial.	22
<b>Tabla 3.</b> Virus de la papa encontrados principalmente en Latinoamérica.	23
<b>Tabla 4.</b> Frecuencia para los sitios de corte simulados en TEV.	95
<b>Tabla 5.</b> Resultados de las frecuencias de corte y mediciones de distancia $S_{\gamma}$ -C y ángulo diedro obtenidas en el sistema PVY.	107
<b>Tabla 6.</b> Resultados de las simulaciones moleculares para PVA.	118

## INTRODUCCIÓN

La familia *Potyviridae* está conformada por virus de plantas de estructura filamentosa y flexuosa con un genoma de RNA de cadena sencilla y sentido positivo. Esta familia está integrada por seis géneros: *Potyvirus*, *Ipomovirus*, *Macluravirus*, *Rymovirus*, *Tritimovirus* y *Bymovirus*. *Potyvirus* es actualmente el grupo más estudiado y numeroso de virus de plantas, con 143 especies reportadas y 32 especies tentativas (King *et al.*, 2012). Los virus de este grupo pueden ser transmitidos por áfidos, de forma mecánica o mediante propagación de material vegetal infectado (e.g. tubérculo-semilla en el caso de la papa) (Petrov *et al.*, 2015).

La distribución geográfica de los potyvirus es muy amplia y su rango de hospedantes incluye los grupos de plantas de mayor importancia económica en el mundo como solanáceas, leguminosas y diversas gramíneas (Scholthof *et al.*, 2011a). Por ejemplo, el *Plum pox virus* (PPV), agente causal de la enfermedad de la *sharka* en cultivos de ciruela (*Prunus* sp.), puede provocar la pérdida total de la producción y se estima que su impacto económico en los últimos 30 años es cercano a 10 billones de dólares (Cambra *et al.*, 2006; Rimbaud *et al.*, 2015). En papaya (*Carica papaya*), la enfermedad causada por *Papaya ringspot virus* (PRSV) figura como la mayor amenaza en la producción de este cultivo tanto en áreas tropicales como subtropicales (Estados Unidos, Suramérica, África, Australia, entre otros), llegando a causar bajas en el rendimiento de hasta el 100% (Tennant *et al.*, 2007; Kalam *et al.*, 2014). Por su parte, *Tobacco etch virus* (TEV) es uno de los principales virus que infectan los cultivos de tabaco (*Nicotiana tabacum*) y otras especies de solanáceas como tomate de mesa (*Solanum lycopersicum*) y pimentón (*Capsicum annuum*), causando necrosis, grabado severo, mosaicos, decaimiento de las plantas y clorosis que resultan en pérdidas del 20 al 70% en la producción anual en Norte y Suramérica (Velásquez *et al.*, 2015). Finalmente, el *Potato virus Y* (PVY) es uno de los virus que generan mayores impactos económicos negativos en diversos cultivos de solanáceas del mundo. Desde el punto de vista taxonómico, PVY está dividido en una compleja clasificación de razas y variantes que pueden inducir desde defoliación y enanismo severo hasta necrosamiento en la superficie de los tubérculos (Beczner *et al.*, 1984; Scholthof *et al.*, 2011b; Karasev y Gray, 2013).

PVY es la especie tipo del género *Potyvirus* y su reporte inicial se remonta a 1930 en investigaciones de los métodos selectivos de transmisión realizadas por Kennet Smith

(Smith, 1931). Entre sus principales hospedantes se encuentran la papa, el tomate de mesa, tomate de árbol (*Solanum betaceum*), tabaco y pimentón (Díaz *et al.*, 2010; Jaramillo *et al.*, 2012; Quenouille *et al.*, 2013a). Este virus es transmitido mecánicamente, por tubérculo-semilla (en papa) y de manera no persistente por más de 100 especies de áfidos. PVY es altamente variable a nivel biológico y molecular gracias a sus elevadas tasas de mutación y recombinación genética, por lo que su clasificación intraespecífica se ha definido con base a cuatro razas principales: PVY<sup>O</sup>, PVY<sup>C</sup>, PVY<sup>Z</sup> y PVY<sup>N</sup>; esta última dividiéndose, a su vez, en las variantes PVY<sup>NTN</sup> y PVY<sup>Wi</sup> (Shukla *et al.*, 1995; Quenouille *et al.*, 2013b; Chikh-Ali y Karasev, 2015). A pesar de ser un virus que se ha estudiado con cierto nivel de profundidad, aún se desconocen aspectos básicos de su ciclo de infección y específicamente de aspectos asociados a la expresión de su genoma (Ivanov *et al.*, 2014). Es sabido que el procesamiento proteolítico de los potyvirus es mediado por las proteasas P1, HC-Pro y NIa-Pro, y constituye un evento crítico en su ciclo infeccioso. NIa-Pro es la más importante de las tres proteasas, al estar involucrada en la maduración de P3, 6K1, CI, 6K2, NIa-Vpg, NIb y CP (Gutiérrez-Campos *et al.*, 1999). Recientemente, en los potyvirus se reportó el descubrimiento de dos nuevas proteínas, PIPO y PISPO, productos de una estrategia novedosa en la expresión de genes, lo que denota lo mucho que resta por conocer sobre la biología molecular de las especies de este género (White, 2015; Mingot *et al.*, 2016).

Al igual que para otras infecciones virales en plantas, no existen agentes químicos de uso práctico en la agricultura para el manejo de enfermedades inducidas por potyvirus, por lo que el manejo técnico de sus vectores resulta fundamental (Díaz *et al.*, 2010), al igual que lo es, el uso de semilla certificada libre de virus, lo que ha promovido estrictos programas de indexación en diferentes países (Haltermann *et al.*, 2012). Debido a la ocurrencia de las variantes necróticas antes indicadas, dichos estándares de certificación han necesitado de técnicas complementarias como pruebas serológicas y métodos moleculares, lo que ha hecho de la detección de virus un proceso complejo y en constante mejoramiento (Singh *et al.*, 2013).

En nuestro país existe un gran desconocimiento de la diversidad de especies virales que infectan cultivos de importancia económica. Para PVY no se cuenta con mucha información relevante respecto a su genoma y/o sintomatología específica que induce en las variedades locales de papa como Diacol-Capiro (*S. tuberosum*), Criolla Colombia (*S. phureja*) o en otras

solanáceas de origen andino como el tomate de árbol, la uchuva (*Physalis peruviana*) y el lulo (*S. quitoense*), entre otros (Jaramillo *et al.*, 2011). Entre los trabajos pioneros se destaca el reporte de la secuencia de un aislamiento de PVY en conjunto con *Potato yellow vein virus* (PVV) en plantas de *Solanum phureja* en Cundinamarca (Villamil *et al.*, 2014) y el registro de variantes filogenéticas de PVY no reportadas en otras partes del mundo (Hena-Díaz *et al.*, 2013). Recientemente, el genoma completo de una cepa de PVY fue reportado por primera vez infectando tomate de forma natural en América, lo que corrobora su amplio rango de hospederos (Muñoz-Baena *et al.*, 2016). Igualmente, la detección de PVY se ha logrado mediante TAS-ELISA y RT-qPCR en una alta proporción de muestras de tubérculos analizados para Criolla-Colombia y Diacol-Capiro, enfatizando la necesidad de utilizar tubérculo-semilla certificado en los cultivos de papa colombianos (Medina *et al.*, 2015).

En este trabajo se planteó caracterizar a partir de resultados de Secuenciación de Nueva Generación (NGS), el genoma completo de variantes de PVY presentes en las regiones cultivadoras de papa de Antioquia, como punto de partida para adelantar estudios de incidencia y diversidad genética de esta especie viral. Adicionalmente, el genoma ensamblado fue utilizado para la obtención de modelos estructurales como base para el establecimiento de un método computacional que permita entender, a escala molecular, el proceso de maduración de la poliproteína de PVY mediado por la proteasa NIa-Pro.

## **Nota al lector**

Atendiendo la normatividad para la presentación de tesis del posgrado en Biotecnología de la Universidad Nacional de Colombia Sede Medellín, dos de los capítulos se presentan directamente como artículos científicos publicados en revistas reconocidas por Colciencias; por lo tanto éstos presentan el formato propio de la respectiva revista.

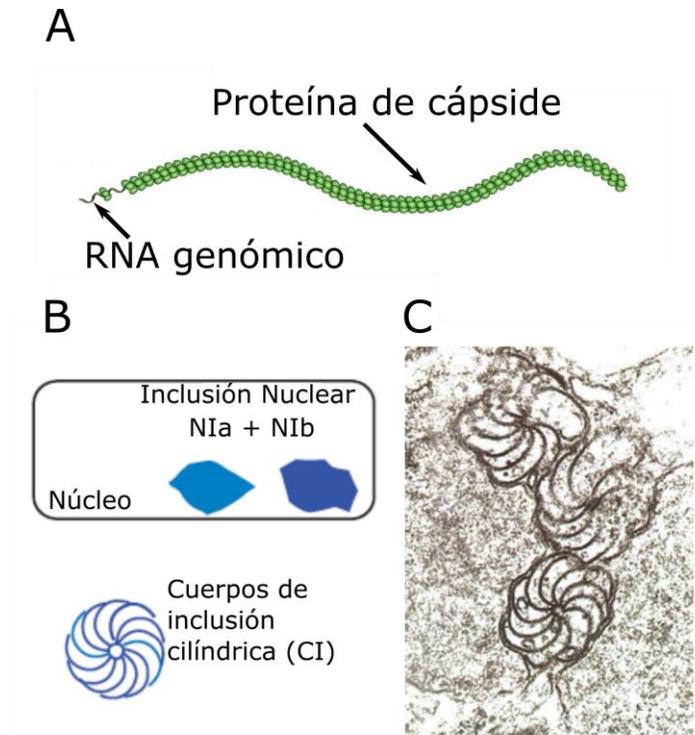
## 1. MARCO TEÓRICO

### 1.1. *Potato virus Y* (PVY)

El género *Potyvirus* conforma el grupo más numeroso de la familia *Potyviridae* y está caracterizado por partículas virales flexuosas y filamentosas sin envoltura externa (King *et al.*, 2012). El genoma de los potyvirus consiste en una molécula de RNA de cadena sencilla y polaridad positiva (ssRNA+), ligada covalentemente a una proteína (VPg) en el extremo 5' y con una cola de poliadenina (poli-A) en el extremo 3'. *Potato virus Y* (PVY) es la especie tipo del género *Potyvirus* y es además una de las que genera mayor impacto en la agricultura mundial. El descubrimiento de PVY proviene de las investigaciones de Kenneth Smith en 1931 sobre la virosis en papa. Gracias al uso de plantas indicadoras y mecanismos selectivos de transmisión, Smith llegó a la conclusión que muchas de las enfermedades presentes en la papa eran causadas por dos especies de virus diferentes a las que llamó: virus X y virus Y, siendo el primero transmitido mecánicamente mientras el segundo podía hacerlo a través del áfido *Myzus persicae*. Posteriormente, estas dos especies fueron nombradas como *Potato virus X* y *Potato virus Y*, respectivamente (Smith, 1931; Mahy y Van Regenmortel, 2010).

El genoma de PVY tiene una longitud aproximada de 9.7 kb y contiene un único ORF que codifica para una poliproteína que es procesada postraduccionalmente en diez polipéptidos funcionales (Revers y García, 2015). Su proteína de cápside (267 aa) es codificada por el fragmento C-terminal de la poliproteína y se encarga de proteger el RNA viral en partículas helicoidales y flexuosas de aproximadamente 740 nm de largo y 11 nm de diámetro (Karasev y Gray, 2013). En 1958, los viroplasmos de PVY fueron observados por primera vez en hojas de *Datura metel* (Trompeta del diablo), al apreciarse patrones cilíndricos de longitud similar organizados en el citoplasma de las células infectadas. Dichas inclusiones cilíndricas fueron igualmente encontradas para otros potyvirus como *Tulip breaking virus* (TBV), *Turnip mosaic virus* (TuMV), *Tobacco etch virus* (TEV), *Wheat streak mosaic virus* (WSMV), *Bean yellow mosaic virus* (BYMV), *Watermelon mosaic virus* (WMV), *Lettuce mosaic virus* (LMV), *Bean common mosaic virus* (BCMV) y *Sugarcane mosaic virus* (SCMV) (David-Ferreira y Borges, 1958; Edwardson, 1966). Estas estructuras de aspás son una de las características

taxonómicas que caracterizan los miembros del género *Potyvirus* y fueron utilizadas por mucho tiempo como herramientas diagnósticas a partir de microscopía óptica (Figura 1-C) (Sorel *et al.*, 2014).



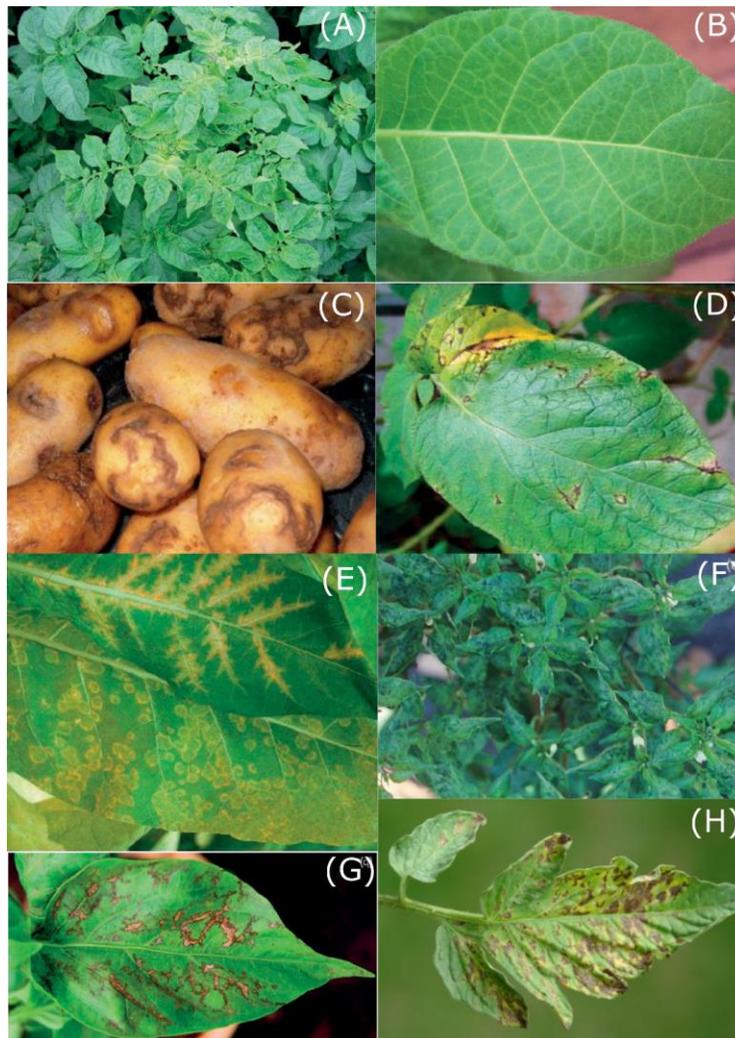
**Figura 1. Características del genoma de PVY. A)** Esquema de la forma de varilla flexuosa del PVY. **B)** Formación de cuerpos de inclusión fuera del núcleo celular que corresponden a **C)** las estructuras únicas en forma de aspas de molino características de miembros del género *Potyvirus* (Tomado de ViralZone; Mahy y Van Regenmortel, 2010).

El rango natural de hospederos para PVY comprende cerca de nueve familias de plantas en las que se incluyen cultivos de importancia económica como pimentón (*Capsicum anuum*), tabaco (*Nicotiana tabacum*), papa (*Solanum tuberosum*), tomate (*Solanum lycopersicum*), berenjena (*Solanum melongena*) y plantas ornamentales como petunia (*Petunia spp.*), entre otras (Jeffries, 1998; Karasev y Gray, 2013). La sintomatología que genera PVY en estas especies consiste generalmente en mosaicos, decaimiento y necrosamiento de diferentes tejidos de las plantas infectadas. Uno de los principales factores que afecta el desarrollo de síntomas inducidos por PVY en papa es la cantidad de variantes virales que aparecen, frecuentemente, como resultado de la agricultura moderna en donde nuevos cultivos susceptibles reemplazan a las variedades tradicionales tolerantes, incrementando las

posibilidades de recombinación genética entre cepas virales (Salazar *et al.*, 1988; Moury, 2010).

Los virus económicamente importantes, causan alteraciones a nivel de tallos y hojas, donde es frecuente observar cambios en el color, manifestadas a través de aclaramiento de venas, mosaicos y amarillamientos (Castro y Contreras, 2011a; Barragán y Guzmán-Barney, 2014). En cuanto a la textura y forma, pueden presentarse hojas pequeñas, enrollamiento, encrespamiento y deformación foliar. En etapas avanzadas de la infección también es factible observar necrosis foliar, que a su vez puede clasificarse de tres formas: apical, si ocurre en los ápices principales y laterales de las plantas; sistémica, cuando son distribuidos en todo el follaje; y necrosis de venas, que ocurre directamente sobre los haces vasculares de las hojas. En los tubérculos también es posible encontrar síntomas que alteren su forma y textura, entre ellos se tienen: alteraciones de la sección transversal normalmente redondeada del tubérculo (forma ahusada), abultamiento de las yemas, rajaduras, flacidez y necrosis tanto superficial (patrones en forma de anillo que se forman en la cáscara) como interna (manchas o puntos circulares dentro del tubérculo) (Castro y Contreras, 2011b). Cabe destacar que la necrosis es uno de los síntomas causados en papa por cepas de al menos once especies virales diferentes y hace de los tubérculos un producto poco comercializable para consumo fresco o para procesamiento por la industria de alimentos (Gil *et al.*, 2011).

En tabaco, PVY modifica el contenido de nicotina en las hojas curadas de la planta, desencadenando grandes pérdidas económicas (Latorre *et al.*, 1984; Doubnerová *et al.*, 2009). En tomate y pimentón, se ha visto la generación de necrosis en hojas (Mascia *et al.*, 2010) (Figura 2). PVY también se ha reportado causando infecciones mixtas con otros virus como *Cucumber mosaic virus* (CMV) y *Potato virus X* (PVX; *Potexvirus*) (Pacheco *et al.*, 2012).

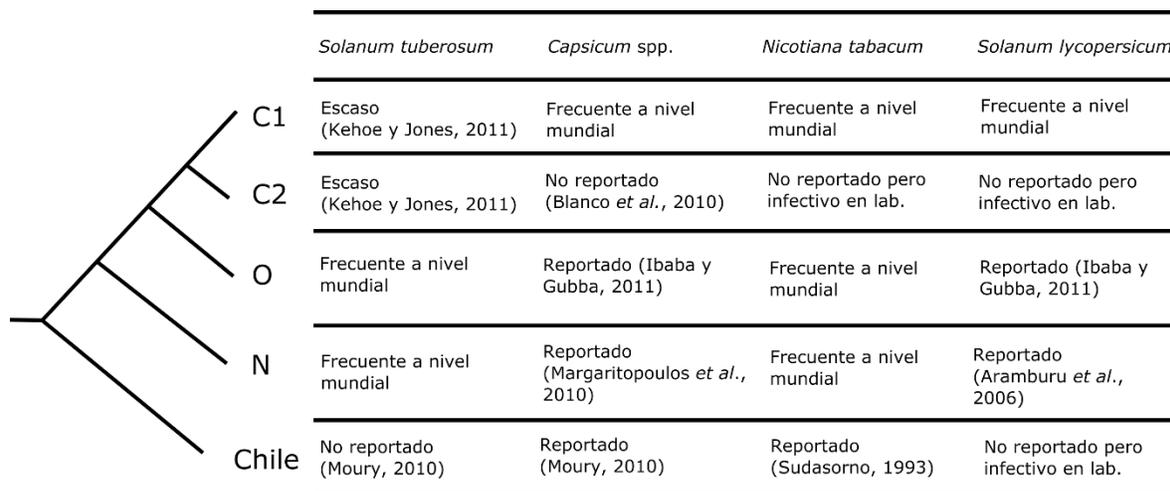


**Figura 2. Principales síntomas inducidos por PVY. (A)** Reducción en el tamaño, amarillamiento y rugosidad foliar en cultivos de papa. **(B)** Clareamiento de venas en una hoja de tabaco. **(C)** sintomatología asociada a *Potato tuber necrotic ringspot disease* (PTNRD) en tubérculo y en **(D)** hoja basal (amarillamiento y necrosamiento). **(E)** Infección de PVY en tabaco cv *Burley*. **(F)** Arrugamiento severo y deformación foliar en hojas de pimentón. **(G)** Patrones necróticos en hojas de pimentón. **(H)** Arrugamiento suave, distorsión y moteado suave en tomate (Tomado de Mahy y Van Regenmortel, 2010).

### 1.1.1. Diversidad del *Potato virus Y*

Filogenéticamente, PVY pertenece a un grupo de 19 potyvirus denominado clado PVY dentro del género *Potyvirus* (Moury y Verdin, 2012). Al interior de esta agrupación, se encuentran 16 especies en América (de las cuales 11 son exclusivas de este continente), sugiriendo que el centro de origen y diversificación del clado PVY está en el continente americano, probablemente en Suramérica. El análisis de genomas ha revelado cinco clados principales

dentro de la especie PVY: C1, N, O, C2 y Chile; siendo los dos últimos los más restringidos en su distribución geográfica y de rango de hospederos (Figura 3).



**Figura 3. Filogenia simplificada de PVY.** Se incluyen las razas y/o variantes más importantes en la infección de cultivos de la familia *Solanaceae* (las variantes recombinantes no se presentan en este gráfico). Nótese que el grupo de Chile presentó la primera diversificación durante la evolución del virus (Adaptado de Quenouille *et al.*, 2013).

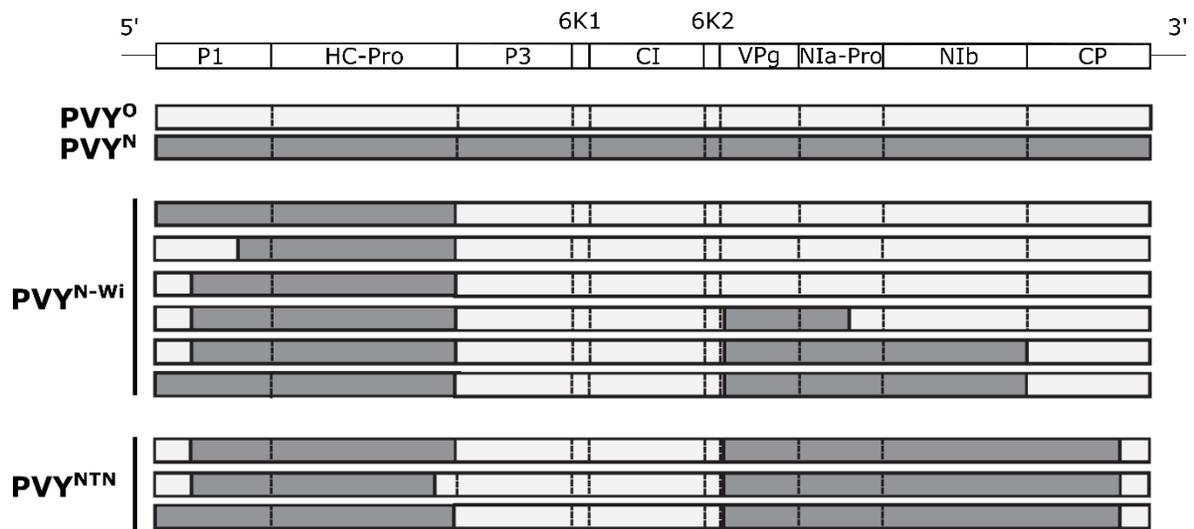
Los aislamientos de PVY en papa han sido distinguidos según criterios patogénicos y moleculares y tienen una compleja clasificación debido a su diversidad biológica, serológica y molecular (Mahy y Van Regenmortel, 2008). El primer nivel de clasificación consta de tres razas principales: PVY<sup>O</sup>, PVY<sup>N</sup>, PVY<sup>C</sup>. Históricamente, estos grupos fueron discriminados por la respuesta hipersensible (HR) que inducen en cultivares de papa que albergaban el gen N, y por los síntomas que causan en plantas de tabaco (Lacroix *et al.* 2011). Específicamente, la raza PVY<sup>C</sup> comprende aislamientos de PVY que inducen HR en cultivos que presentan el gen N<sub>C<sub>tbr</sub></sub>, mientras que PVY<sup>O</sup> está conformado por aislamientos que realizan lo propio para el gen N<sub>Y<sub>tbr</sub></sub>. Por su parte, las cepas de PVY<sup>N</sup> no inducen HR en los cultivares que presentan estos genes. Sintomáticamente PVY<sup>N</sup> también puede ser discriminado de sus pares al inducir necrosis de venas en tabaco (PVY<sup>C</sup> y PVY<sup>O</sup> inducen mosaico y amarillamiento de venas, solamente). En 1980, un nuevo grupo de cepas denominado PVY<sup>Z</sup> fue hallado al superar los genes N<sub>C</sub> y N<sub>Y</sub>, inducir HR en ciertos cultivares de papa pero no causar necrosamiento de venas (a diferencia de PVY<sup>N</sup>), por lo que se pensó en la presencia de otro gen putativo, N<sub>Z<sub>tbr</sub></sub>. Una década después, otro grupo denominado PVY<sup>E</sup> mostró superar todos los genes de resistencia (N<sub>C</sub>, N<sub>Y</sub> y N<sub>Z</sub>) pero no inducir necrosis en venas (Karasev y Gray, 2013). La Tabla

1 resume la sintomatología y respuestas que causan cada una de las variantes descritas según las características de los cultivares donde fueron identificadas.

**Tabla 1. Respuestas sintomáticas y resistencia hipersensible según las variantes de PVY.** HR: Respuesta hipersensible; Mos: mosaico; s: susceptible; VN: necrosis de venas. (Tomado de Karasev y Gray, 2013).

<b>Característica de los cultivares</b>	<b>PVY<sup>O</sup></b>	<b>PVY<sup>C</sup></b>	<b>PVY<sup>Z</sup></b>	<b>PVY<sup>N</sup></b>	<b>PVY<sup>E</sup></b>
Ny (Desiree, Pentland Crown)	HR	s	s	s	s
Nc (King Edward)	s	HR	s	s	s
Ny, Nc, Nz (Maris Bard, Pentland Ivory)	HR	HR	HR	s	s
Tabaco	Mos	Mos	Mos	VN	Mos

Gracias a los avances en las técnicas de secuenciación aplicados a la virología de plantas, fue posible esclarecer que muchos aislamientos de PVY presentan eventos de recombinación entre las cepas de PVY<sup>O</sup> y PVY<sup>N</sup> (Singh *et al.*, 2008). PVY<sup>N</sup> presenta, a su vez, dos variantes relacionadas y denominadas PVY<sup>NTN</sup> y PVY<sup>W</sup>, siendo el último un genotipo particular producto de eventos de recombinación entre variantes y altas tasas de mutación que induce necrosis de venas en tabaco pero que es serológicamente similar a la raza PVY<sup>O</sup> (razón por la cual en Norteamérica es conocido como PVY<sup>N:O</sup>) (Henao-Díaz *et al.*, 2013). PVY<sup>N:O</sup> es un recombinante entre dos genomas parentales de PVY<sup>N</sup> y PVY<sup>O</sup>; mientras que PVY<sup>NTN</sup>, probablemente provino de la variante PVY<sup>N:O</sup> mediante eventos adicionales de recombinación (Karasev *et al.*, 2011). En la Figura 4 se ilustran los patrones de recombinación entre dos razas de PVY que generan las cepas PVY<sup>N-Wi</sup> y PVY<sup>NTN</sup>.



**Figura 4. Patrones de recombinación entre dos razas de PVY.** Regiones de PVY<sup>O</sup> (gris claro) y PVY<sup>N</sup> (gris oscuro) que son compartidos entre dos variantes principales. En la parte superior se muestra el esquema del genoma de un potyvirus (tomado de Quenoille, 2013).

## 1.2. Enfermedades del cultivo de la papa

Las condiciones ecológicas del cultivo de la papa tienen gran influencia sobre las especies de fitopatógenos a las que es susceptible. Si bien hay más de 300 plagas y enfermedades reportadas afectando este cultivo, no todas causan grandes pérdidas en sus rendimientos. Entre los patógenos más limitantes se destacan el oomycete *Phytophthora infestans* que causa la gota de la papa (Fry *et al.*, 2015), la bacteria *Ralstonia solanacearum* (la marchitez bacteriana) y el protozoario *Spongospora subterranea* f.sp. *subterranea* (la sarna polvosa de la papa) (Osorio *et al.*, 2012). Los nemátodos más limitantes corresponden a *Globodera* sp. y *Meloidogyne* sp.; mientras que los insectos de mayor importancia son la polilla pequeña (*Phthorimaea operculella*) y la polilla grande de la papa (*Tecia solanivora*), generalmente controlados con plaguicidas químicos (Castro y Contreras, 2011). Los áfidos, también se incluyen en este conjunto por ser vectores de muchos virus (Petrov *et al.*, 2015). En la Figura 5 se ilustran los microorganismos causales de algunas de las enfermedades más importantes en papa.



**Figura 5. Microorganismos asociados a las principales enfermedades en papa no causadas por virus. A)** *Phytophthora infestans* responsable del tizón tardío o gota de la papa. (Tomado de Fry y Grünwald, 2010). **B)** El nemátodo *Meloidogyne* sp. causa agallas irregulares en la raíz de la papa. (Tomado de Lazo, 2010; Bayer, 2016). **C)** El protozoario *Spongospora subterranea* asociado a la sarna polvosa de la papa (Tomado de Unimedios, 2008; Balendres *et al.*, 2017).

Por otra parte, se han reportado cerca de 38 especies de virus que afectan el cultivo de la papa, algunos de los cuales dependen de este hospedante para su diseminación: *Potato virus S* (PVS), *Potato virus M* (PVM), *Potato leafroll virus* (PLRV), *Potato mop-top virus* (PMTV), *Andean potato latent virus* (APLV). Otros virus presentan otros hospederos principales como: *Cucumber mosaic virus* (CMV), *Tobacco ringspot virus* (TRSV), *Tobacco rattle virus* (TRV), *Tobacco mosaic virus* (TMV) y *Alfalfa mosaic virus* (AMV) (Valkonen *et al.*, 1994; Al-Saleh *et al.*, 2014). La infección combinada con diversas especies virales puede desencadenar en una serie de afecciones en el crecimiento de las plantas que termina en grandes pérdidas (más del 90 %) en el rendimiento de los cultivos (Roossinck, 2015; Rodríguez *et al.*, 2015). Tal reducción es medida en variables como el tamaño, número,

aspecto y contenido de tubérculos, y cambian según la especie, variante y mecanismo de transmisión del virus en cuestión (Figura 6). La Tabla 2 resume los ocho virus más importantes en el mundo para este cultivo (Mahy y Van Regenmortel, 2008).



**Figura 6. Síntomas inducidos por diversas especies virales en papa.** **A)** Infección secundaria por PLRV en donde se aprecia retraso en el crecimiento y enrollamiento foliar. **B)** Arcos necróticos en tubérculos asociado con TRV. **C)** Síntomas de PMTV en tubérculos de papa. **D)** Necrosis localizada inducida por PLRV en tubérculos de papa (Tomado de Mahy y Van Regenmortel, 2008).

**Tabla 2. Principales virus reportados en cultivos de papa a nivel mundial.** NP: No persistente; P: Persistente. (Adaptado de Mahy y Van Regenmortel, 2008).

Especie y año del primer reporte	Género	Importancia económica	Distribución geográfica	Transmisión	Otros hospederos	Variabilidad
<i>Potato virus Y</i> (PVY) 1931	<i>Potyvirus</i>	Muy alta	Mundial	Áfidos (NP)	Tabaco, tomate, pimentón	Cepas: PVY <sup>O,N,C</sup> Variantes: PVY <sup>NTN,N-W,N:O</sup>

<i>Potato Leafroll virus</i> (PLRV) 1916	<i>Polerovirus</i>	Muy alta	Mundial	Áfidos (P)	<i>Solanum</i> spp., tomate	Baja variabilidad
<i>Potato virus X</i> (PVX) 1931	<i>Potexvirus</i>	Alta	Mundial	Contacto	Tomate, pimentón	Cepas 1, 2, 3, 4 Serotipos PVX <sup>O</sup> , PVX <sup>A</sup>
<i>Potato virus A</i> (PVA) 1932	<i>Potyvirus</i>	Moderada	Mundial	Áfidos (NP)	Sólo papa	Cuatro cepas Tres serotipos
<i>Potato virus S</i> (PVS) 1952	<i>Carlavirus</i>	Moderada	Mundial	Áfidos (NP) Contacto	<i>Solanum</i> spp.	Cepas PVS <sup>O</sup> , PVS <sup>A</sup>
<i>Potato virus M</i> (PVM) 1923	<i>Carlavirus</i>	Moderada	Mundial	Áfidos (NP) Contacto	Solanáceas mayormente	Cepa PVM-ID
<i>Tobacco rattle virus</i> (TRV) 1946	<i>Tobravirus</i>	Alta	Mundial	Nemátodos	Tabaco, lechuga, espinaca	Diversas cepas
<i>Potato mop-top virus</i> (PMTV) 1966	<i>Pomovirus</i>	Moderada	Climas fríos	pseudohongo	Solo papa	PMTV-S, PMTV-T

En las últimas cuatro décadas, se ha registrado un incremento en el número de virus reportados que afectan el cultivo de papa en América Latina, especialmente en la región Andina (Quenouille *et al.*, 2013). Un estudio realizado por Kerlan *et al.* (2008) registró la incidencia de cerca de 16 virus encontrados principalmente en Latinoamérica cuya importancia económica y mecanismos de transmisión aún no se hallan completamente esclarecidos; la Tabla 3 describe algunas de estas especies.

**Tabla 3. Virus de la papa encontrados principalmente en Latinoamérica.** NP: transmisión no persistente. SP: transmisión semipersistente.

Virus	Género	Distribución geográfica	Transmisión	Referencia
<i>Andean potato latent virus</i>	<i>Tymovirus</i>	Bolivia, Colombia, Ecuador, Perú	Contacto, semilla, coleópteros	(Kreuze <i>et al.</i> , 2009)

(APLV)				
<i>Andean potato mottle virus</i> (APMoV)	<i>Comovirus</i>	Chile, Ecuador, Perú, Brazil	Contacto, semilla, coleópteros	(Neves-Borges et al., 2001)
<i>Potato virus V</i> (PVV)	<i>Potyvirus</i>	Colombia	Áfidos (NP)	(Gutiérrez et al., 2014)
<i>Potato virus T</i> (PVT)	<i>Tepovirus</i>	Perú, Bolivia	Mecánica, semilla	(Hooker 1981)
<i>Tobacco ringspot virus</i> (TRSV)	<i>Nepovirus</i>	Perú	Nemátodos, semilla	(Hooker 1981)
<i>Potato yellow vein virus</i> (PYVV)	<i>Crinivirus</i>	Colombia, Ecuador, Perú	Mosca blanca (SP), tubérculos	(Rodríguez et al., 2015)

En conjunto con *Potato virus A* y *Potato leaf roll virus* (PLRV; *Polerovirus*), PVY representa una de las mayores amenazas para la producción del cultivo de papa en el mundo, generando pérdidas en el rendimiento de hasta un 90% (Salazar, 2003). Entre los diez principales virus de plantas, PVY ha sido catalogado en el puesto número cinco debido a la ausencia de fuentes de resistencia eficiente en los cultivares de papa frente a las variantes del virus que inducen síntomas necróticos tanto en hojas como tubérculos (Scholthof et al., 2011a). El desconocimiento de las interacciones PVY-hospedero y falencias en las herramientas disponibles para su diagnóstico han llevado a que las razas necróticas de PVY sean responsables de pérdidas económicas valoradas en cientos de millones de dólares al año sólo en Norteamérica (Kogovsek et al., 2008; Scholthof et al., 2011b).

Además del tubérculo-semilla, en papa la diseminación de los virus ocurre de una manera altamente eficiente, gracias a diversos vectores artrópodos. Los áfidos son el grupo más importante de vectores de virus de plantas. Pertenecen a la familia *Aphididae* y rara vez causan daños a las plantas durante sus procesos de alimentación. La manera cómo se comportan los áfidos sobre las plantas, así como su ciclo de vida y las condiciones ambientales, determinan el tipo y la cantidad de virus que puedan transmitir (Quenouille et al., 2013). *Myzus persicae* es uno de los vectores más importantes y eficientes de PVY y prácticamente se encuentra en todos los lugares donde se cultiva papa; otros ejemplos son: *Aphis fabae*, *Cryptomyzus galeopsidis* y *Hyalopterus erysimi* (Verbeek et al., 2010). Existen tres tipos de clasificación de los mecanismos de transmisión asociados a estos vectores: persistente, semipersistente y no persistente (Maxwell et al., 2016). *Potyvirus* es el género más numeroso y conocido que se transmite de manera no persistente, siendo ejemplos típicos PVA y PVY; pero también otras especies como PVS y PVM (*Carlavirus*) presentan

este tipo de transmisión. En este sistema de transmisión no existe un período de latencia, es decir que la adquisición e inoculación del virus se llevan a cabo en segundos o pocos minutos (Leclerc *et al.*, 2014). Los áfidos pierden la capacidad de transferir el virus una vez ocurre la muda de los estiletes viejos, la faringe y la membrana que recubre el intestino anterior (Quenouille *et al.*, 2013). Recientemente, se demostró cómo una adquisición preferencial de la variante PVY<sup>NTN</sup> sobre PVY<sup>O</sup> por *M. persicae*, se debía a la expresión del gen de resistencia N<sub>Y</sub> en el cultivo que confería una resistencia parcial a la variante ordinaria (Carroll *et al.*, 2016). Adicionalmente, el trabajo de Govier y Kassanis (1974) demostró la dependencia de proteínas para la transmisión no persistente de los virus. En el caso de los potyvirus, trabajos posteriores han demostrado que la transmisión es dependiente de la proteína no estructural HC-Pro (*Helper-Component Protease*) que lo asiste en su proceso infectivo (Sasaya *et al.*, 2000; Syller, 2006).

### **1.2.1. Cultivo de papa en Colombia: Incidencia de PVY**

El cultivo de papa en Colombia es uno de los renglones agrícolas más importantes de la zona altoandina. En el año 2000, este cultivo alcanzó 2.8 millones de toneladas sembradas en 163.841 ha y un rendimiento de 17.3 t/ha; muy por encima al de países vecinos como Bolivia, Perú y Ecuador. En el 2007, el área del cultivo se vio reducida a 110 mil ha, con una producción de 1.9 millones de toneladas (FAO, 2008). En 2013, Cundinamarca, Boyacá y Nariño, fueron los principales departamentos productores de papa con 19.57, 15.37 y 9.78%, de la producción, respectivamente; con un área nacional de cultivo de 151616 ha, 2.7 millones de toneladas y un rendimiento de 18.4 t/ha (Agronet, 2016). Adicionalmente, el cultivo de la papa tiene un efecto significativo sobre otros renglones de la economía regional al requerir una elevada cantidad de plaguicidas químicos y fertilizantes (en segundo lugar después del café) y generar la demanda más alta de transporte terrestre en el país (Toro *et al.*, 2013). En el ámbito social, más de 500 mil personas son dependientes directas del cultivo de la papa, además de comerciantes, transportadores, procesadores y productores. Dada su condición de especie eje en los sistemas de producción de clima frío, representa una fuente de ingresos y generación de empleos directos a nivel rural para más de 95 mil familias en el país (Cuartas, 2006).

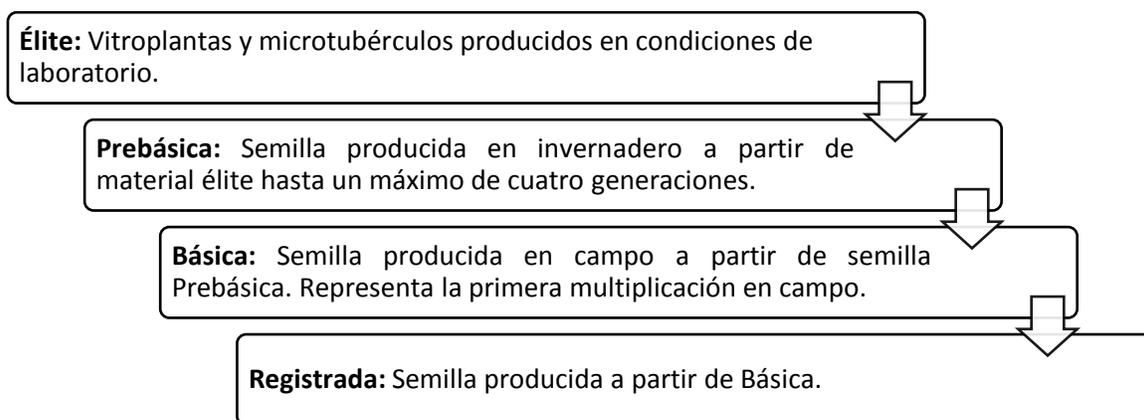
La información de los niveles de incidencia, severidad y efecto de PVY en los cultivos de papa de Colombia es escasa, al punto que se desconocen las características genotípicas de sus principales variantes en el país. En 1980 se reportaron pérdidas del 61% en el rendimiento de la variedad de papa Puracé como resultado de una infección mixta de PVY con PVX y PLRV (Guerrero-Guerrero y Martínez-López, 1980). Sánchez y colaboradores evaluaron la incidencia de PVY en coinfección con PVX, PVS y PLRV mediante técnicas serológicas en tres zonas agroecológicas de cultivo (3200 msnm, 2600 msnm y 2150 msnm), obteniendo como resultado una incidencia del 87% en la zona baja, siendo la mayor entre las tres y causando una disminución en el rendimiento del 40%. A esto, los autores atribuyeron la alta población de áfidos como la principal causa (Sánchez *et al.*, 1991). Gil y colaboradores reportaron en 2011 niveles de incidencia del 72% mediante pruebas de ELISA y secuenciación de cepas colombianas de PVY en departamentos como Antioquia, Boyacá, Cundinamarca y Nariño (Gil *et al.*, 2011). Adicionalmente, mostraron una asociación de las cepas estudiadas con la raza necrosante PVY<sup>N</sup> y su variante PVY<sup>NTN</sup>. Otro estudio realizado por Henao-Díaz y colaboradores en estos departamentos contribuyó con un análisis filogenético para 24 aislamientos de PVY en cultivos de papa y tomate de árbol, donde se encontraron tres variantes principales del virus, dos de las cuales (I-Col, IV-Col) no se asociaron con ninguna secuencia de referencia a las razas tradicionales; mientras que la tercera hizo parte del clado PVY<sup>NTN</sup> (Henao-Díaz *et al.*, 2013). En otras investigaciones más recientes se mostró una incidencia del 46% en plantas con infección mixta de PVV y PVY que en principio parecía relacionarse con una sintomatología atípica en papa. Los resultados de la investigación no permitieron concluir que dicha coinfección era causal de los nuevos síntomas observados, por lo que se dejó abierta la posibilidad de la presencia de un virus no reportado en los cultivos de papa bajo estudio (Alfaro y Franco-Lara, 2015). Finalmente, un estudio realizado en tubérculos para *S. tuberosum* y *S. phureja* reveló que la detección de PVY se ve comprometida cuando se utilizan pruebas de menor sensibilidad al obtener incidencias del 75% y 0% mediante TAS-ELISA, respecto a cifras del 100% y 66.66% con RT-qPCR, respectivamente (Medina *et al.*, 2015).

La secuenciación de nueva generación (NGS) representa otra metodología en el campo de la virología vegetal, implementándose desde el año 2009 gracias al trabajo de Kreuze y colaboradores para identificar y ensamblar el genoma completo de *Sweet potato feathery mottle virus* (SPFMV) en plantas asintomáticas (Kreuze *et al.*, 2009). Gracias a esta técnica

se han logrado diagnosticar virus en diferentes plantas hospedantes, identificar nuevas especies y obtener sus genomas completos (Boonham *et al.*, 2014; Khodakov *et al.*, 2016). En Colombia, diferentes estudios han permitido la identificación y reporte de virus infectando cultivos de importancia económica utilizando NGS. En *S. phureja* se logró obtener el genoma completo de PVS en tejido foliar en Antioquia, siendo el primero reportado para la región Andina suramericana (Gutiérrez *et al.*, 2013). Similarmente, en uchuva se detectó la infección de PVX que permitió ensamblar su genoma, conformando la única secuencia completa reportada de un potexvirus en Colombia (Gutiérrez *et al.*, 2015). Otras investigaciones reportaron la coinfección de PYVV, PVV y PVY en muestras de *S. tuberosum* en Cundinamarca usando NGS y confirmación por RT-PCR (Villamil-Garzón *et al.*, 2014). Recientemente, se obtuvo la secuencia completa de PVY infectando tomate en Antioquia, siendo el primer reporte en América del virus para este hospedero (Muñoz-Baena *et al.*, 2016).

### **1.3. Métodos de manejo de enfermedades virales**

Dado que las enfermedades virales no pueden ser manejadas con métodos curativos como en el caso de las causadas por bacterias u hongos, se recurre entonces a métodos preventivos que dependen en gran medida del conocimiento de los factores ecológicos y biológicos que determinen la prevalencia de esas enfermedades en el cultivo (Díaz *et al.*, 2010). En papa, el método tradicional de control es la producción de semilla asexual libre de virus y de otros agentes patógenos. Aquellas denominadas como líneas élite (procedentes de plantas *in vitro*) y semilla pre-básica (de producción en invernadero) son la parte inicial del proceso (Figura 7). En las etapas de campo (ej. básica, registrada, certificada) existe una constante verificación de la totalidad de las plantas en búsqueda de virus mediante pruebas de detección como ELISA y RT-PCR (Halterman, 2012).



**Figura 7. Diagrama de generaciones en la producción de semilla de papa.** Etapas del proceso de semilla certificada en el cultivo de papa.

En Norteamérica, el cultivo de papa tiene un sistema de certificación de semillas riguroso y eficiente (Gray *et al.*, 2010a). Este sistema ha permitido a países como Estados Unidos y Canadá ser más conscientes del espectro de síntomas causados por diversas variantes de PVY. Sin embargo, es claro que la presencia del virus en esta región consiste de un complejo de razas con una preponderancia de aislamientos pertenecientes a PVY<sup>O</sup>, lo que hace difícil la identificación de ciertos síntomas suaves característicos de las razas PVY<sup>N</sup> y PVY<sup>N-Wi</sup> (Mondal *et al.*, 2017). Dado que algunos de estos síntomas pueden pasar por alto en las inspecciones visuales propias de sus sistemas de certificación, se emplean entonces técnicas de diagnóstico molecular como pruebas de ELISA y RT-PCR en formato multiplex, que permiten la identificación cepas e inclusive, de infecciones mixtas (Damirdagh y Ross, 1967; Nie y Singh, 2002; Gray *et al.*, 2010b).

En Europa, debido a que PVY<sup>NTN</sup> ha desplazado notoriamente la predominancia de la raza PVY<sup>O</sup> en los últimos 30 años, también se emplea un conjunto de pruebas biológicas, serológicas y moleculares para identificar efectivamente esta y otras variantes (Rolland *et al.*, 2008; Visser *et al.*, 2012). En países como Inglaterra los límites permitidos en semillas certificadas no superan niveles del 0.1% de plantas con síntomas virales para semilla prebásica; 0.8% para semilla básica y 6% para semilla certificada (DEFRA, 2016). En Estados Unidos esta última categoría permite niveles no superiores al 5% (Schramm, 2011).

Los virus transmitidos por áfidos constituyen un grave problema en sistemas de cultivo abierto. Dentro de las medidas disponibles para remediar esta situación se destaca el uso de barreras vivas, las cuales interfieren con la actividad del vector y actúan como sumidero de

virus transmitidos de manera no persistente, pues la incidencia del virus puede verse reducida cuando una proporción de los vectores pierden su carga infectiva en el cultivo barrera antes de llegar al cultivo susceptible (Avilla *et al.* 1996). En estudios orientados a analizar la diseminación de PVY se hace uso de una malla fina (20/10 mesh) con doble puerta que recubre las dimensiones del cultivo. Con base en ello, se ha logrado apreciar que la incidencia de virus no superó el 20% bajo la malla anti-áfido, en contraste con un 100% alcanzado en un escenario al aire libre. Dicha exclusión permitió además aumentar el rendimiento del cultivo en un 37.8% (Sepúlveda *et al.*, 2009). Sin embargo, investigaciones recientes apuntan al fortalecimiento de los cultivares de papa enfocándose en la siembra de materiales mejorados por sus resistencia a virus (Valkonen, 2015).

Otros métodos que ayudan a reforzar el control de enfermedades virales abarcan las prácticas culturales como la limpieza de herramientas y la rotación de cultivos. El descarte, o raleo, durante el cultivo se realiza eliminando las plantas sintomáticas lo más temprano posible, evitando que sirvan como fuente de inóculo. Luego del cultivo, el descarte ocurre por eliminación de tubérculos con síntomas de enfermedad (USDA, 2016).

#### **1.4. Biología molecular de PVY**

Como se mencionó anteriormente, PVY tiene un genoma monopartita de RNA de cadena positiva de cerca de 9.7 kb . Este genoma codifica para un marco de lectura que produce una poliproteína de 340 – 370 kDa que es procesada postraduccionalmente por las proteasas virales P1, HC-Pro y NIa-Pro en diez proteínas maduras. Las proteínas P1 y HC-Pro están ubicadas en el extremo N-terminal y son liberadas autocatalíticamente, probablemente justo después de ser sintetizadas (King *et al.*, 2012). Las demás proteínas (P3, 6K1, CI, 6K2, NIa-VPg, NIa-Pro, NIb y CP) son cortadas a partir del precursor de alto peso molecular mediante la acción de la cisteín-proteasa NIa-Pro (Adams *et al.*, 2005). Tanto las proteínas maduras como algunos intermediarios de este procesamiento tienen funciones relevantes en el ciclo infectivo de los potyvirus, los cuales parecen realizarse de una manera coordinada e interdependiente (Figura 8).

P3 es una de las proteínas menos caracterizadas de los potyvirus. Ha sido asociada con las inclusiones cilíndricas citoplasmáticas formadas por la proteína CI, al igual que por las inclusiones nucleares de NIa y NIb (Langenberg y Zhang, 1997). Recientemente, la región hidrofóbica de su extremo C-terminal fue identificada como responsable de la interacción con las membranas del retículo endoplasmático y por permitir la formación de vesículas replicativas (Cui *et al.*, 2010). Adicionalmente, su presencia es requerida para la replicación viral y es relevante en términos de patogenicidad y sintomatología. Su separación del precursor P3-6K1 no es esencial en la infectividad del virus; más aún, probablemente en dicho intermediario recaen las funciones principales y su procesamiento proteolítico suceda como mecanismo de regulación (Riechmann *et al.*, 1995). No obstante, en PPV se demostró que 6K1 existe como proteína madura individual (6 kDa) y, aunque no queda claramente definida su participación y función en la multiplicación del virus, se sabe que el corte entre estas dos proteínas afecta la expresión de genomas, por lo que 6K1 puede ser relevante en la infección potyviral (Waltermann y Maiss, 2006).

CI es una proteína multifuncional que forma inclusiones cilíndricas en forma de aspas de molino en el citoplasma de células infectadas. Presenta actividades ATPasa y RNA helicasa, requeridas en la replicación del RNA viral (Fernández *et al.*, 1997). Se ha encontrado su asociación con el movimiento de translocación del virus a través de plasmodesmas y en la descapsidación de la partícula infecciosa. Adicionalmente, se ha demostrado que actúa como factor de virulencia para diversos genes de resistencia de la planta (ej. eIF4E, PSI-K) (Sorel *et al.*, 2014). La formación del complejo replicativo (VRC) es iniciada por el precursor 6K2-VPg, cuyo primer componente elicita a NIa para interactuar con NIb en el retículo endoplasmático, formando un complejo de replicación asociado a la membrana (Ivanov *et al.*, 2014). También, CI-6K2 es parcialmente responsable por el reclutamiento de componentes necesarios para que dicho complejo inicie la síntesis de (-)RNA. Esto se fundamenta por la presencia de CI-6K2 durante mayores intervalos de tiempo que CI-6K2-NIa, sugiriendo que el sitio de corte del primero es procesado mucho más lento (Merits *et al.*, 2002).

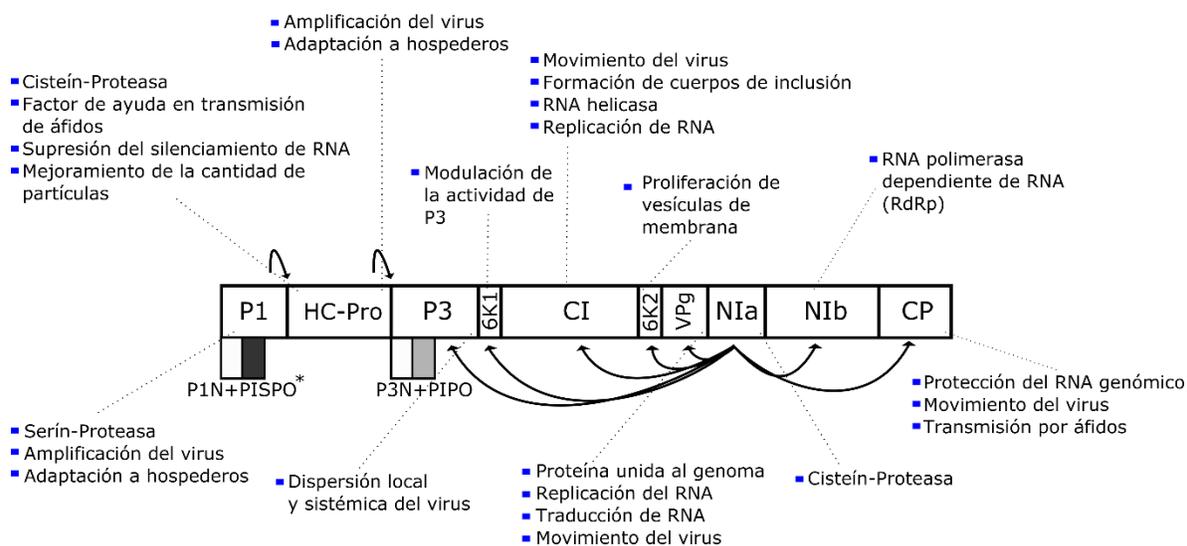
NIa es la proteína más grande y puede formar inclusiones tanto en el núcleo como en el citoplasma de las células. Es parcialmente procesada para dar lugar a VPg y a su dominio proteasa, NIa-Pro (Dougherty y Dawn, 1991; Martín *et al.*, 1992). VPg presenta una característica de flexibilidad necesaria para interacciones con diferentes proteínas en

diversos procesos. En la mayoría de los casos se encuentra unida en el extremo 5' del RNA genómico; aunque el precursor VPg-NIa-Pro también ha sido detectado hacia el final del mismo (Mathur y Savithri, 2012a). En asociación con la proteína 6K2 (precursor 6K2-VPg-NIa-Pro), VPg es dirigida hacia las membranas donde cumple un papel importante en la replicación del RNA del virus. El sitio de corte entre el dominio VPg y la proteasa es subóptimo debido a la presencia del glutamato en lugar de la glutamina de la posición P-1 (residuo donde se efectúa la hidrólisis del enlace peptídico); se cree que el procesamiento de VPg es un evento crítico para la replicación viral (Mathur y Savithri, 2012b). VPg puede ser modificada postraduccionalmente mediante fosforilación y se ha demostrado que presenta un mayor número de interacciones con proteínas del virus y el hospedante (Bosque *et al.*, 2014).

NIb corresponde a la RNA polimerasa dependiente de RNA (o RNA replicasa) y es la encargada de la replicación del genoma en los potyvirus (Hong y Hunt, 1996). Su acción se desarrolla en los complejos replicativos ubicados en las estructuras membranosas en conjunto con algunos factores del hospedero (eEF1A, PABP) y con el precursor 6K2-VPg-NIa-Pro que ayuda a su direccionamiento (Dufresne *et al.*, 2008). Su actividad inicia una vez emplea a la proteína VPg como cebador para dar paso a la síntesis del RNA viral y su presencia en el núcleo muy seguramente está asociada a un mecanismo de regulación (Revers y García, 2015).

Finalmente, la función principal de la proteína CP reside en la encapsidación del genoma viral. Se requieren cerca de 2000 subunidades de cápside arregladas helicoidalmente alrededor del RNA genómico para dar origen a un virión flexuoso de 11 a 13 nm de diámetro y 680 a 900 nm de largo (King *et al.*, 2012). La región central es altamente conservada y sus extremos N y C-terminal son esenciales para el ensamblaje entre subunidades, siendo la primera muy variable (Rybicki y Shukla, 1992; Ksenofontov *et al.*, 2013). En PVA y PPV han sido reportadas modificaciones postraduccionales en CP que se atribuyen como elementos de control para regular las fracciones de RNA genómico que se dirigen a procesos de traducción, replicación o propagación según las etapas del proceso de infección (Ivanov *et al.*, 2003; Šubr *et al.*, 2010). A parte de la encapsidación, CP también ha sido involucrada en la amplificación del genoma, movimiento y transmisión (Revers y García, 2015).

Recientemente, fueron descubiertas en los potyvirus dos nuevas proteínas denominadas PISPO y PIPO, las cuáles son generadas por ORF alternativos como resultado de deslizamientos en el marco de lectura de la polimerasa viral en P1 y P3, respectivamente. El descubrimiento de P3N-PIPO, representa una estrategia novedosa en la expresión de genes de virus de plantas (White, 2015). También es interesante la supresión del silenciamiento de RNA encontrado en *Sweet potato feathery mottle* (SPFMV) llevada a cabo por P1N-PISPO, característica que hasta el momento era atribuída sólo a HC-Pro (Mingot *et al.*, 2016).



**Figura 8. Características relevantes de la poliproteína potyviral.** Las flechas indican los sitios de corte por las tres proteasas: P1, HC-Pro y NIa-Pro. PIPO es representado en color gris claro debajo del ORF complementario en P3. (Adaptado de Revers y García, 2015). \*Hasta el momento, la proteína PISPO en la región P1 ha sido reportada exclusivamente en *Sweet potato feathery mottle virus* (SPFMV) (Mingot *et al.*, 2016).

#### 1.4.1. Procesamiento de la poliproteína en PVY

El procesamiento proteolítico de una poliproteína es una estrategia de traducción común durante el ciclo de infección de muchos virus, que permite un uso eficiente de la capacidad de codificación del genoma. En PVY, y en potyvirus en general, el procesamiento de la poliproteína es llevado a cabo por la acción de las proteasas P1, HC-Pro y NIa-Pro. La proteína P1 es la más variable en tamaño y secuencia (30-63 kDa) en los potyvirus, particularmente en su extremo N-terminal; la región C-terminal de P1 tiene un dominio serín-proteasa que se encarga de realizar el corte entre las proteínas P1 y HC-Pro (Adams *et al.*,

2005; Yoshida *et al.*, 2012). El sitio catalítico de P1 está determinado por la presencia de la serina, cuyo grupo hidroxilo es el responsable de causar el ataque nucleofílico que genera la hidrólisis del enlace peptídico (al igual que la tripsina y la quimotripsina). Se ha reportado que P1 no es esencial para la viabilidad del virus pero sí es necesaria su separación de la proteína HC-Pro (Pasin *et al.*, 2014), además ayuda a estimular la amplificación del genoma y tiene la capacidad de interactuar con ácidos nucleicos *in vivo*, por lo que se ha pensado en su probable participación en la supresión del silenciamiento de RNA (Valli *et al.*, 2007).

La segunda proteasa, conocida como HC-Pro (*Helper Component Protease*), es una cisteín proteasa multifuncional asociada a la supresión de genes y a la transmisión por vectores, relacionándose con los mecanismos de defensa del virus. Puede agregarse en forma de inclusiones citoplasmáticas y parece cumplir una función estabilizadora para CP en los extremos de los viriones (Valli *et al.*, 2014). HC-Pro tiene tres dominios funcionales, siendo el CPD (dominio cisteína proteasa) en la región C-terminal el que le permite liberarse de la proteína mayor mediante un corte en el sitio HC-Pro/P3 (Syller, 2006; Guo *et al.*, 2011), mientras que su extremo N-terminal está asociado a la transmisión por áfidos (Govier y Kassanis, 1977; Plisson *et al.*, 2003). También se ha demostrado que está asociada con otras proteínas como CI, P1, VPg y NIa-Pro; aunque la relevancia de estas interacciones no esté clara todavía (Bosque *et al.*, 2014).

Por último, NIa-Pro es responsable del procesamiento proteolítico de las regiones centrales y C-terminales de la poliproteína (Adams *et al.*, 2005). NIa-Pro resulta clave en la expresión del genoma en potyvirus ya que es responsable de los siete sitios de corte restantes en la poliproteína (Dougherty *et al.*, 1988a; Yoon *et al.*, 1999). Esta proteasa comparte motivos con serín proteasas como la quimotripsina, pero su sitio activo tiene una cisteína en lugar de una serina que, junto con la histidina y el ácido aspártico, conforman la tríada catalítica (Kapust *et al.*, 2001). La NIa-Pro del TEV, es la más estudiada entre todos los potyvirus. En el año 2002 logró resolverse su estructura por cristalografía en asociación con el sustrato y el producto, lo que permitió detallar los requisitos estructurales que definen su estricta especificidad de secuencia (Dougherty *et al.*, 1988b; Phan *et al.*, 2002). Los estudios sobre la maduración postraduccional de potyvirus concluyen que, muy probablemente, estos deben controlar su procesamiento proteolítico con el fin de optimizar el proceso de infección y replicación. Lo anterior puede llevarse a cabo mediante variaciones en las secuencias blanco de NIa-Pro,

permitiendo la aparición de las proteínas maduras en la parte del ciclo de infección donde ejercen su función principal (Yoon *et al.*, 1999; Merits *et al.*, 2002). Esto ha sido respaldado por el hecho que es posible incrementar la resistencia a los potyvirus a través de plantas transgénicas que expresan inhibidores de proteasas (Gutierrez-Campos *et al.*, 1999; Gholizadeh *et al.*, 2005). Lo anterior indica que es factible entender aspectos fundamentales sobre el ciclo de vida de potyvirus mediante el estudio detallado del mecanismo de corte de la proteasa NIa-Pro. Para tal propósito pueden utilizarse modelos computacionales basados en la estructura cristalográfica de la NIa-Pro de TEV, aprovechando la conservación de secuencias entre este tipo de proteasas para estudios similares sobre otros potyvirus. Uno de los principales objetivos de este trabajo consiste en explorar la viabilidad de usar métodos de simulación molecular para el estudio de la proteasa NIa-Pro utilizando la información de secuencia obtenida a partir de estudios de secuenciación masiva y predecir la cinética de procesamiento de la poliproteína mediada por esta proteasa.

### **1.5. Métodos de Dinámica Molecular en los estudios de virus**

La simulación por Dinámica Molecular (MD) es una técnica computacional que permite simular el cambio en la posición de un conjunto de átomos en función del tiempo a una temperatura y presión determinada (Frenkel, 2002; Karplus y McCammon, 2002). Con el progreso de la capacidad computacional, los métodos de Dinámica Molecular cada vez pueden abarcar sistemas de mayor tamaño con simulaciones que pueden llegar incluso a escalas de milisegundos (Ode *et al.*, 2012). Los métodos por MD han extendido rápidamente la gama de aplicaciones en biología, ayudando a revelar características únicas de estructuras de proteínas que serían difíciles de obtener por métodos experimentales (Ulf, 2008). En virología, la mayoría de las investigaciones que emplean MD se han centrado en comprender la integridad estructural, estabilidad y conformación de las proteínas de cápside en conjunto con ácidos nucleicos (DNA y RNA). Otras simulaciones muestran cómo la estabilidad de las partículas virales se ve influenciada por una fuerte interacción entre los componentes lipídicos de los virus con envoltura (Reddy, 2016). También han ayudado a comprender aspectos como la respuesta del virus frente a los sistemas de defensa del hospedero, la estructura y función de las enzimas virales y viriones, los mecanismos de resistencia a drogas y al descubrimiento y desarrollo de compuestos antivirales (Ode *et al.*, 2012; Perilla

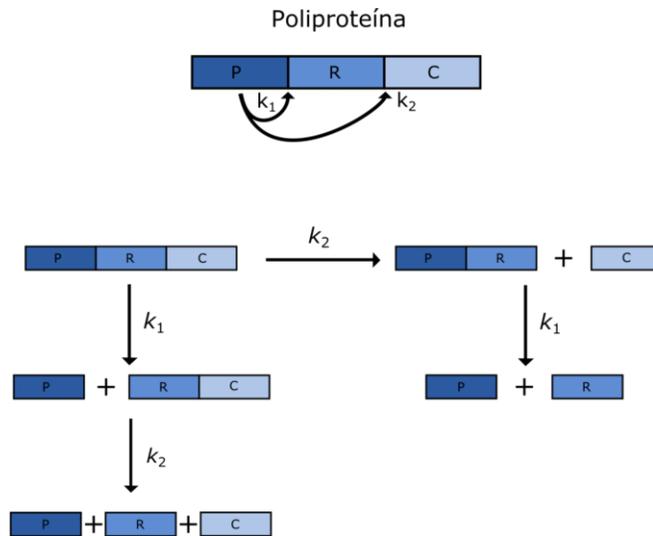
*et al.*, 2016). Por ejemplo, utilizando simulaciones por MD se demostró cómo la región N-terminal expuesta al solvente del péptido de fusión de HIV-1 (esencial en el proceso infeccioso), puede ser un potencial objetivo para el diseño de vacunas (Kong *et al.*, 2016). La predicción de sitios de unión de flavonoides en el proceso de entrada a la célula de DENV2 (*Dengue virus*), ejemplifica otras de las aplicaciones de la simulación molecular al evaluar la afinidad de la interacción por los puentes de hidrógeno de éstos compuestos con la proteína E del virus (Ismail y Jusoh, 2017). Otro ejemplo se presenta en el virus de la hepatitis C (HCV), de donde se han estudiado dos importantes zonas antigénicas de la glicoproteína E2, aportando estrategias terapéuticas alternativas a las ya existentes (Akbar y Jusoh, 2013).

La utilidad de las simulaciones MD no se limita a encontrar posibles blancos terapéuticos analizando la conformación de sus estructuras proteicas. Diversas investigaciones se han dirigido a analizar etapas fundamentales del ciclo de vida del virus como el reconocimiento de unión entre proteína-sustrato durante el procesamiento proteolítico. Recientemente, en el virus del dengue, se llevó a cabo una serie de simulaciones computacionales para su proteasa NS2B/NS3 en complejo con seis tipos de sustrato con el fin de comparar la especificidad del reconocimiento de unión para cada uno. Como resultados se obtuvieron diferencias de unión en diferentes subsitios, que dependían de la estabilidad del sitio activo según la conformación de algunos residuos de su proteasa (Yotmanee *et al.*, 2015). En experimentos de mutagénesis dirigida se ha demostrado que al cambiar la región N-terminal de la proteína potyviral (cambio de Gln por Leu o Pro) queda abolida la capacidad replicativa de un clon infeccioso de TEV en protoplastos (Adams *et al.*, 2005). Otras investigaciones afines aportan información sobre el uso de métodos para el cálculo de energías libres de unión (Hu *et al.*, 2016); o la determinación de los estados de protonación de residuos en el diseño de ligandos (Gueto-Tettay, *et al.*, 2016).

Lo anterior sugiere que la Dinámica Molecular podría ser utilizada como un método para la estimación de las tasas relativas de corte en el procesamiento proteolítico de cada uno de los siete blancos de la proteasa NIa-Pro en la poliproteína viral. La estimación de estas tasas de corte podría ser utilizada posteriormente en modelos cinéticos para predecir los tiempos de maduración de cada proteína viral y plantear hipótesis sobre el ciclo de vida de diferentes potyvirus. Como se mencionó anteriormente, la dinámica del proceso postraduccional de potyvirus tiene implicaciones en la regulación de la aparición consecutiva y acumulación de

intermediarios de las proteínas maduras finales (Mahy y Van Regenmortel, 2008). A modo de ejemplo, en la Figura 10 se ilustran las ecuaciones cinéticas de una reacción proteolítica para un virus hipotético cuyo mecanismo de expresión ocurre a través de una poliproteína que madura en tres proteínas individuales: P (proteasa), R (replicasa) y C (cápside). En este modelo, la poliproteína es procesada por el dominio P través de los sitios de reconocimiento P/R y R/C, con constantes cinéticas  $k_1$  y  $k_2$ , respectivamente.

Este sistema puede ser modelado mediante seis ecuaciones diferenciales que representan las tasas de formación y desaparición de cada una de las especies involucradas y de los productos finales esperados. Inicialmente, se observa que la poliproteína PRC presenta una única tasa de desaparición (no hay un precursor que genere PRC, Figura 9-a) que se realiza mediante dos procesos: un evento intramolecular definido por una constante de reacción  $k_1$ , y un evento intermolecular determinado por la constante  $k_2$ . En el primer caso, la proteasa se corta autocatalíticamente lo que permite obtener el dipéptido RC a una velocidad proporcional a la concentración de la poliproteína y a la constante  $k_1$  (Figura 9-b). Adicionalmente, este precursor presenta una tasa de desaparición que ocurre cuando la proteasa actúa entre la región de corte de la replicasa y la cápside, en un proceso intermolecular proporcional a  $k_2$ . El segundo caso posible en el proceso de maduración de la poliproteína PRC involucra la liberación de la cápside cuando la proteasa y replicasa aún se encuentran unidas en un mismo dominio. Por lo tanto, la tasa de aparición de la especie PR es proporcional a la constante del proceso intermolecular,  $k_2$ ; una vez que la proteasa se libera intramolecularmente, desaparecerá con una velocidad proporcional a  $k_1$  (Figura 9-c). De lo anterior se concluye que la producción de la proteasa como una proteína madura depende exclusivamente de un proceso autocatalítico proporcional a  $k_1$  y de las concentraciones de las especies PRC y PR (Figura 9-d). Por su parte, la replicasa puede ser obtenida cuando el intermediario PR es procesado unimolecularmente y cuando el dipéptido RC se separa mediante un proceso bimolecular (Figura 9-e). Finalmente, la cápside sólo puede ser obtenida mediante eventos intermoleculares proporcionales a las concentraciones de la poliproteína y del intermediario RC.

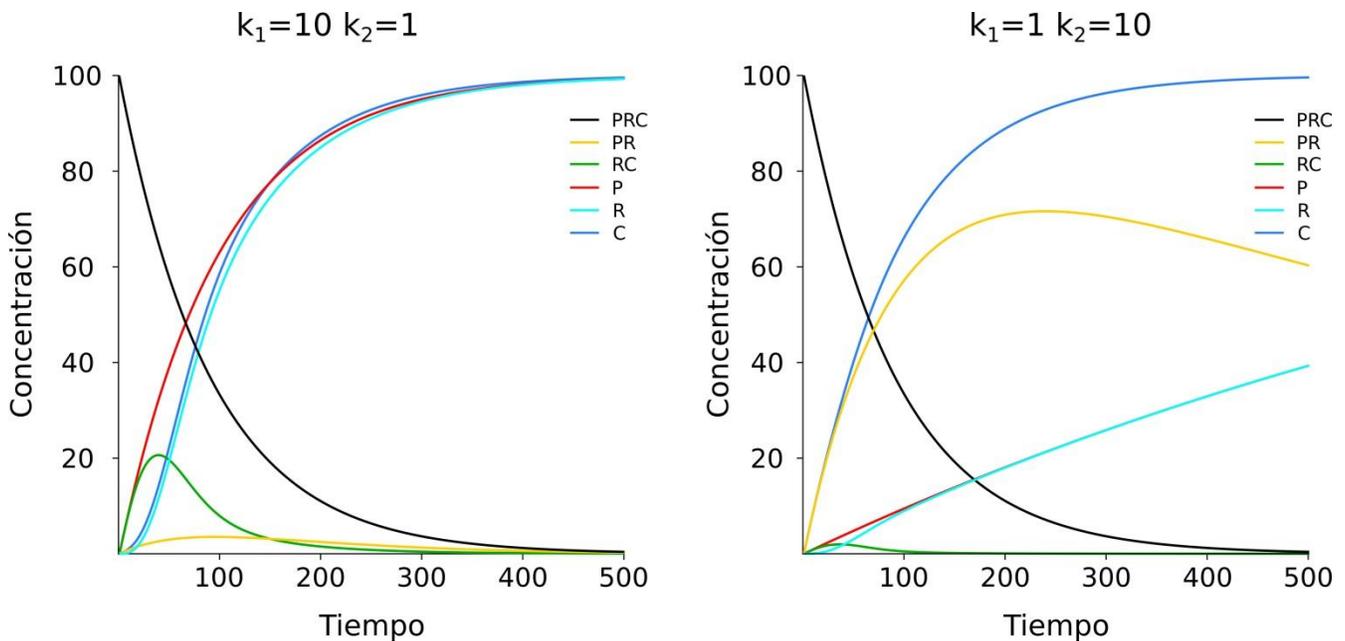


$$\begin{array}{ll}
 \text{a) } \frac{d\text{PRC}}{dt} = -(k_1+k_2)[\text{PRC}] & \text{d) } \frac{d\text{P}}{dt} = (k_1)[\text{PRC}]+(k_1)[\text{PR}] \\
 \text{b) } \frac{d\text{RC}}{dt} = (k_1)[\text{PRC}]-k_2[\text{RC}] & \text{e) } \frac{d\text{R}}{dt} = (k_1)[\text{PR}]+(k_2)[\text{RC}][\text{P}] \\
 \text{c) } \frac{d\text{PR}}{dt} = (k_2)[\text{PRC}]-k_1[\text{PR}] & \text{f) } \frac{d\text{C}}{dt} = (k_2)[\text{PRC}]+(k_2)[\text{P}][\text{RC}]
 \end{array}$$

**Figura 9. Modelo de maduración enzimática para un virus de tres componentes.** Las ecuaciones describen la tasa a la que desaparecen las especies intermediarias luego del corte realizado por la proteasa. También se muestran las tasas de formación de las proteínas maduras, una vez son procesados los dos precursores RC y PR.

La magnitud de las constantes cinéticas en el proceso de maduración influye de manera distinta sobre la evolución de cada una de las especies. Como se observa en la Figura 10, un valor diez veces mayor para  $k_1$  sobre  $k_2$  establece una cinética más eficiente que en el caso contrario. La principal diferencia consiste en que la producción de las proteínas maduras ocurre en la mitad del tiempo establecido y aproximadamente en tasas de velocidad similares. Esto es coherente con el procesamiento de la poliproteína de la Figura 9, pues entre más rápido se genere la separación autocatalítica de la proteasa, habrá una mayor concentración disponible para procesar el intermediario RC y así obtener la replicasa y la cápside. Adicionalmente, la alta tasa de desaparición del precursor PR se evidencia en la poca concentración que esta especie alcanza y en el corto tiempo que permanece durante la etapa de desaparición del complejo PRC. Por otro lado, al reducir la magnitud de  $k_1$  se observa que la concentración del precursor PR resulta mucho mayor en comparación al caso anterior. El bajo procesamiento del evento intramolecular no permite obtener la concentración

suficiente de la proteasa para separar eficientemente el intermediario RC. La cinética de las especies se reduce considerablemente excepto por la cápside que muestra el mismo perfil de formación anterior. Esto se explica porque se favorece su tasa de separación de la poliproteína. A partir de lo anterior puede inferirse que el ciclo de infección del virus hipotético está regido por el proceso de maduración que ejecuta sobre su poliproteína.



**Figura 10. Efecto de las constantes de afinidad sobre la cinética de maduración de un virus hipotético con una poliproteína de tres componentes P/R/C (P: proteasa; R: replicasa y C: cápside).** En el panel izquierdo se muestra el efecto de tener una valor para la constante del proceso intramolecular en relación 10:1 respecto al evento intermolecular. En este caso todas las proteínas maduran aproximadamente al mismo tiempo (300 unidades de tiempo) por lo que el proceso es más eficiente. En el panel derecho los sitios de corte P/R y R/C presentan valores de constantes cinéticas en relación 1:10. Para este escenario, ocurre una rápida aparición de la cápside respecto a la replicasa, cuya concentración no aumenta. Escala de tiempo en unidades relativas.

Por lo tanto, para estudiar la cinética de las proteínas maduras de potyvirus mediada por Nla-Pro es necesario estimar los valores relativos para la eficiencia de corte en cada uno de sus siete sitios blanco. Además, este modelo debe incluir un sistema de 36 ecuaciones diferenciales donde se representan cada uno de los posibles intermediarios y proteínas finales. En el primer capítulo de este trabajo se estudia la viabilidad de usar simulaciones MD para la predicción de tasas relativas de corte para cada uno de los sitios reconocidos por Nla-Pro y que puedan ser incorporadas en un modelo de predicción cinética sobre la

maduración de los potyvirus. Este modelo es utilizado para realizar predicciones sobre aislamientos de PVY en Antioquia que fueron secuenciados completamente por métodos de NGS, como se describe en los artículos correspondientes a los capítulos 4.2 y 4.3.

## 2. BIBLIOGRAFÍA

Adams M, Antoniw J, Beaudoin F. Overview and analysis of the polyprotein cleavage sites in the family *Potyviridae*. *Molecular Plant Pathology*. 2005; 6:471-487.

Agronet, Ministerio de Agricultura y desarrollo rural. 2016. Disponible en: <http://www.agronet.gov.co/Paginas/estadisticas.aspx> (Acceso mayo 2016).

Akbar R, Jusoh S. Stability, orientation and position preference of the stem region (residues 689-703) in *Hepatitis C Virus* (HCV) envelope glycoprotein E2: a molecular dynamics study. 2013; 2:1-12.

Alfaro J, Franco-Lara L. *Potato Virus Y* (PVY) y *Potato Yellow Vein Virus* (PYVV) en Infecciones Mixtas No Causan Síntomas Atípicos en Plantas de Papa. *Revista Facultad de Ciencias Básicas*. 2015; 11:26-37.

Al-Saleh M, Amer M, Al-Shahwan I, Abdalla O, Shakeel M. Molecular characterization of Two *Alfalfa Mosaic Virus* Isolates Infecting Potato Crop in Central Region of Saudi Arabia. *Internatinal Journal of Agriculture and Biology*. 2014; 16:976-980.

Avilla C, Collar J, Duque M, Hernáiz P, Martín B, Ferreres A. Cultivos barrera como método de control de virus no persistentes en pimiento. *Boletín de sanidad vegetal. Plagas*. 1996; 22:301-307.

Balendres M, Nichols D, Tegg R, Wilson C. Potato Root Exudation and Release of *Spongospora subterranea* Resting Spore Germination Stimulants are Affected by Plant and Environmental Conditions. *Journal of Phytopathology*. 2017; 165:64-72

Bayer. Nemátodo de la raíz. Crop Science Chile. 2016. Disponible en: <http://cropscience.bayer.cl/soluciones/fichaproblema.asp?id=1073> (Acceso mayo 2016).

Beczner L, Horváth J, Romhányi I, Förster H. Studies on the etiology of tuber necrotic ringspot disease in potato. *Potato Research*. 1984; 27:339-352.

Boonham N, Kreuze J, Winter S, van der Vlugt R, Bergervoet J, Tomlinson J, Mumford R. Methods in virus diagnostics: from ELISA to next generation sequencing. *Virus Research*. 2014; 186:20-31.

Bosque G, Folch-Fortuny A, Picó J, Ferrer A, Elena S. Topology analysis and visualization of *Potyvirus* protein-protein interaction network. *BMC Systems Biology*. 2014; 8:1-15.

Cambra M, Capote N, Myrta A, Llácer G. *Plum pox virus* and the estimated costs associated with sharka disease. *Bulletin OEPP/EPPO*. 2006; 36:202-204.

Carroll J, Smith D, Gray S. Preferential acquisition and inoculation of PVY<sup>NTN</sup> over PVY<sup>O</sup> in potato by the green peach aphid *Myzus persicae* (Sulzer). *The Journal of general virology*. 2016; 97:797-802.

Castro I, Contreras A. Manejo de plagas y enfermedades en el cultivo de la papa. 1 ed. Valdivia: Editorial Imprenta Austral; 2011. p. 40-50.

Chikh-Ali M, Karasev A. Immunocapture-Multiplex RT-PCR for the Simultaneous Detection and Identification of Plant Viruses and Their Strains: Study Case, *Potato Virus Y* (PVY). *Methods in Molecular Biology*. 2015; 1302:177-186.

Cuartas P. Estudio del desarrollo de la infección y del efecto de la formulación de aislamientos de granulovirus sobre *Tecia solanica* y *Phthorimaea operculella* (Trabajo de grado). Bogotá: Microbiología Agrícola y Veterinaria, Facultad de Ciencias Básicas, Pontificia Universidad Javeriana. 2006:3-7.

Cui X, Wei T, Chowda-Reddy R, Sun G, Wang A. The *Tobacco etch virus* P3 protein forms mobile inclusions via the early secretory pathway and traffics along actin microfilaments. *Virology*. 2010; 397:56-63.

Damirdagh I, Ross A. A marked synergistic interaction of potato viruses X and Y in inoculated leaves of tobacco. *Virology*. 1967; 31:296-307.

David-Ferreira J, Borges M. Virus na célula vegetal observada em lâminas ao microscópio eletrônico 1-virus Y da Batata. *Boletim de Sociedade Brasileira de Botânica*. 1958; 32:329-332.

DEFRA. Agriculture in the UK. Defra Statistics, York, UK. 2016. Disponible en <https://www.gov.uk/government/organisations/department-for-environment-food-rural-affairs/about/statistics> (Acceso junio 2016).

Díaz A, Quiñones M, Arana F, Soto M, Hernández A. *Potyvirus*: Características generales, situación de su diagnóstico y determinación de su presencia en el cultivo de pimiento en Cuba. *Revista de Protección Vegetal*. 2010; 25:69-79.

Doubnerová V, Müller K, Cerovská N, Synková H, Spoustová P, Ryslavá H. Effect of *Potato virus Y* on the NADP-malic enzyme from *Nicotiana tabacum* L.: mRNA, expressed protein and activity. *International Journal of molecular sciences*. 2009; 13:3583-3598.

Dougherty W, Carrington J, Cary S, Parks T. Biochemical and mutational analysis of a plant virus polyprotein cleavage site. *EMBO Journal*. 1988; 7:1281-1287.

Dougherty W, Dawn T. Post-translational processing of the *Tobacco etch virus* 49-kDa small nuclear inclusion polyprotein: Identification of an internal cleavage site and delimitation of VPg and proteinase domains. *Virology*. 1991; 183:449-456.

Dufresne P, Thivierge K, Cotton S, Beauchemin C, Ide C, Ubalijoro E, *et al.* Heat shock 70 protein interaction with *Turnip mosaic virus* RNA-dependent RNA polymerase within virus-induced membrane vesicles. *Virology*. 2008; 374:217-227.

Edwardson J. Electron microscopy of cytoplasmic inclusions in cells infected with rod shapes viruses. *American Journal of Botany*. 1966; 53:359-364.

FAO. International year of the potato. 2008. Disponible en: <http://www.fao.org/potato-2008/en/world/latinamerica.html> (Acceso mayo 2016).

Fernández A, Guo H, Sáenz P, Simón-Buela L, Gómez de Cedrón M, García J. The motif V of plum pox potyvirus CI RNA helicase is involved in NTP hydrolysis and is essential for virus RNA replication. *Nucleic Acids Research*. 1997; 25:4474-4480.

Frenkel D, Smit B. *Understanding molecular simulation: from algorithms to applications*. 2 ed. San Diego: Elsevier Academic Press; 2002. p. 63-107.

Fry W, Birch P, Judelson H, Grünwald N, Danies G, Everts K, *et al.* Five Reasons to Consider *Phytophthora infestans* a Reemerging Pathogen. *Pytophatology*. 2015; 105:966-981.

Fry W, Grünwald N. Introduction to Oomycetes. *The Plant Health Instructor*. 2010. DOI:10.1094/PHI-I-2010-1207-01

Gholizadeh A, Santha I, Kohnehrouz B, Lodha M, Kapoor H. Cystatins may confer viral resistance in plants by inhibition of a virus-induced cell death phenomenon in which cysteine proteinases are active: cloning and molecular characterization of a cDNA encoding cysteine-proteinase inhibitor (celostatin) from *Celosia cristata* (crested cock's comb). *Biotechnology and Applied Biochemistry*. 2005; 42:197-204

Gil J, Cotes J, Marín M. Incidencia de potyvirus y caracterización molecular de PVY en regiones productoras de papa (*Solanum tuberosum* L.) de Colombia. *Revista Colombiana de Biotecnología*. 2011; 13:85-93.

- Govier D, Kassanis B. A virus-induced component of Plant Sap needed when aphids acquire *Potato virus Y* from purified preparations. *Virology*. 1974; 61:420-426.
- Govier D, Kassanis B, Pirone T. Partial purification and characterization of the *Potato virus Y* helper component. *Virology*. 1977; 78:306-314.
- Gray S, De Boer S, Lorenzen J, Karasev A, Whitworth J, Nolte P *et al.* *Potato virus Y*: An Evolving Concern for Potato Crops in the United States and Canada. *Plant Disease*. 2010; 94:1384-1396.
- Guerrero-Guerrero O, Martínez-López G. Evaluación de pérdidas ocasionadas en la variedad de papa Puracé por los virus "*Potato virus X*", "*Potato virus Y*" y "*Potato Leafroll virus*". *Fitopatología Colombiana*. 1980; 9:3-40.
- Gueto-Tettay C, Pestana-Nobles R, Drosos-Ramirez J. Determination of the protonation state for the catalytic dyad in  $\beta$ -secretase when bound to hydroxyethylamine transition state analogue inhibitors: A molecular dynamics simulation study. 2016; 66:155-167.
- Guo B, Lin J, Ye K. Structure of the Autocatalytic Cysteine Protease Domain of *Potyvirus* Helper-component Proteinase. *The Journal of Biological Chemistry*. 2011; 286:21937-21943.
- Gutiérrez P, Alzate J, Montoya M. Complete genome sequence of an isolate of *Potato virus X* (PVX) infecting Cape gooseberry (*Physalis peruviana*) in Colombia. *Virus Genes*. 2015; 50:518-522.
- Gutiérrez P, Alzate J, Marín-Montoya M. Complete genome sequence of a novel potato virus S strain infecting *Solanum phureja* in Colombia. *Archives of Virology*. 2013; 61:2205-2208.
- Gutierrez-Campos R, Torres-Acosta J, Saucedo-Arias L, Gómez-Lim M. The use of cysteine proteinase inhibitors to engineer resistance against potyviruses in transgenic tobacco plants. *Nature Biotechnology*. 1999; 17:1223–1226.

Halterman D, Charkowski A, Verchot J. Potato, viruses, and seed certification in the USA to provide healthy propagated tubers. *Pest Technology*. 2012; 6:1-14.

Henao-Díaz E, Gutiérrez-Sánchez P, Marín-Montoya M. Análisis filogenético de aislamientos del *Potato virus Y* (PVY) obtenidos en cultivos de papa (*Solanum tuberosum*) y tomate de árbol (*Solanum betaceum*) en Colombia. *Actualidades Biológicas*. 2013; 35:219-232.

Hong Y, Hunt A. RNA polymerase activity catalyzed by a potyvirus-encoded RNA-dependent RNA polymerase. *Virology*. 1996; 226:146-151.

Hooker W, editor. *Compendium of Potato Diseases*. USA: American Phytopathological Society; 1981; 2:76-100.

Hu G, Ma A, Dou X, Zhao L, Wang J. Computational Studies of a Mechanism for Binding and Drug Resistance in the Wild Type and Four Mutations of HIV-1 Protease with a GRL-0519 Inhibitor. *International journal of molecular sciences*. 2016; 17:2-15.

Ismail N, Jusoh S. Molecular Docking and Molecular Dynamics Simulation Studies to Predict Flavonoid Binding on the Surface of DENV2 E Protein. *Interdisciplinary sciences*. En prensa.

Ivanov K, Eskelin K, Lohmus A, Mäkinen K. Molecular and cellular mechanisms underlying potyvirus infection. *Journal of General Virology*. 2014; 95:1415-1429.

Ivanov K, Puustinen P, Gabrenaite R, Vihinen H, Ronnstrand L, Valmu L *et al.* Phosphorylation of the potyvirus capsid protein by protein kinase CK2 and its relevance for virus infection. *The Plant Cell*. 2003; 15:2124-2139.

Jaramillo M, Álvarez J, Marín M. Características de los virus asociados a la virosis del tomate de árbol (*Solanum betaceum*) en Colombia. *Revista Lasallista de Investigación*. 2012; 9:115-127.

Jaramillo M, Gutiérrez P, Lagos L, Cotes J, Marín M. Detection of a complex of viruses in tamarillo (*Solanum betaceum*) orchards in the Andean region of Colombia. *Tropical Plant Pathology*. 2011; 36:150-159

Jeffries K. Technical Guidelines for the Safe Movement of Germplasm. Food and Agriculture Organization of the United Nations. 1998; 19:70-76.

Kalam A, Amin L, Sidik N. Gene technology for *Papaya Ringspot Virus* Disease Management. *The Scientific World Journal*. 2014; 1:1-11.

Kapust R, Tözsér J, Fox J, Anderson E, Cherry S, Copeland T, *et al.* Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Engineering*. 2001; 14:993-1000.

Karasev A, Gray S. Continuous and Emerging Challenges of *Potato virus Y* in Potato. *Annual review of Phytopathology*. 2013; 51:571-586.

Karasev A, Hu X, Brown C, Kerlan C, Nikolaeva O, Crosslin J, *et al.*, Genetic diversity of the ordinary strain of *Potato virus Y* (PVY) and origin of recombinant PVY strains. *Phytopathology*. 2011; 101:778-785.

Karplus M, McCammon J. Molecular dynamics simulations of biomolecules. *Nature structural biology*. 2002; 9:646-652.

Khodakov D, Wang C, Zhang D. Diagnostics based on nucleic acid sequence variant profiling: PCR, hybridization, and NGS approaches. *Advanced Drug Delivery Reviews*. 2016; 105:3-19.

King A, Adamas M, Carstens E, Lefkowitz E, editores. *Virus Taxonomy: Classification and Nomenclature of Viruses*. US: Elsevier Academic Press. 2012:1072-1080.

Kogovsek P, Gow L, Pompe-Novak M, Gruden K, Foster G, Boonham N, Ravnikar M. Single-step RT real-time PCR for sensitive detection and discrimination of *Potato virus Y* isolates. *Journal of virological methods*. 2008; 149:1-11.

Kong R, Xu K, Zhou T, Acharya P, Lemmin T, Liu K, *et al.* Fusion peptide of HIV-1 as a site of vulnerability to neutralizing antibody. *Science*. 2016; 352:828-833.

Kreuze J, Perez A, Untiveros M, Quispe D, Fuentes S, Barker I, *et al.* Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. *Virology*. 2009; 388:1–7.

Ksenofontov AL, Paalme V, Arutyunyan A, Semenyuk P, Fedorova N, Rumvolt R *et al.* Partially disordered structure in intravirus coat protein of Potyvirus *Potato virus A*. *PLoS One*. 2013; 8:1-7.

Lacroix C, Glais L, Verrier JL, Jacquot E. Effect of passage of a *Potato virus Y* isolate on a line of tobacco containing the recessive resistance gene *va2* on the development of isolates capable of overcoming alleles 0 and 2. *Plant Pathology*. 2011; 130:259–269.

Langenberg W, Zhang L. Immunocytology shows the presence of *Tobacco etch virus* P3 protein in nuclear inclusions. *Journal of Structural Biology*. 1997; 118:243-247.

Latorre B, Flores V, Marholz G. Effect of *Potato Virus Y* on Growth, yield, and chemical composition of flue-cured tobacco in Chile. *Plant Disease*. 1984; 68:884-886.

Lazo M. Diagnóstico de nemátodos en el cultivo de patata (*Solanum tuberosum* L.). 2010. Disponible en: <http://cursa.ihmc.us/rid=1JL7FP2RC-2CBCDQ0-Y6W/1JKNP6FTDI19681RLI4N1limage> (Acceso abril 2016).

Leclerc M, Doré T, Gilligan C, Lucas P, Filipe. Estimating the Delay between Host Infection and Disease (Incubation Period) and Assessing Its Significance to the Epidemiology of Plant Diseases. *PLoS ONE*. 2014; 9:1-12.

Mahy B, Van Regenmortel M, editores. Desk Encyclopedia of Plant and fungal Virology. España: Academic Press. 2010:458-480.

Mahy B, Van Regenmortel M, editores. Encyclopedia of Virology. Eslovenia: Academic Press. 2008:458-471.

Martín M, García J, Cervera M, Goldbach R, van Lent J. Intracellular localization of three non-structural plum pox potyvirus proteins by immunogold labelling. *Virus Research*. 1992; 25:201-211.

Mascia T, Cillo F, Fanelli V, Finetti-Sialer M, De Stradis A, Palukaitis P, *et al.* Characterization of the interactions between *Cucumber mosaic virus* and *Potato virus Y* in mixed infections in tomato. 2010; 23:1514-1524.

Mathur C, Savithri H. Novel ATPase activity of the polyprotein intermediate, viral protein genome-linked-nuclear inclusion-a protease, of *Pepper vein banding potyvirus*. *Biochemical and Biophysical Research Communications*. 2012; 427:113-118.

Maxwell D, Partridge J, Roberts N, Boonham N, Foster G. The Effects of Plant Virus Infection on Polarization Reflection from Leaves. *PLoS One*. 2016; 11:1-18.

Merits A, Rajamäki M, Lindholm P, Runeberg-Roos P, Kekarainen T, Puustinen P, *et al.* Proteolytic processing of potyviral proteins and polyprotein processing intermediates in insect and plant cells. *Journal of General Virology*. 2002; 83:1211-1221.

Medina C, Gutiérrez P, Marín M. Detección del *Potato virus Y* (PVY) en tubérculos de papa mediante TAS-ELISA y qRT-PCR en Antioquia (Colombia). *Bioagro*. 2015; 27:83–92.

Mingot A, Valli A, Rodamilans B, San León D, Baulcombe D, García J, *et al.* The P1N-PISPO trans-Frame Gene of *Sweet Potato Feathery Mottle Potyvirus* Is Produced during Virus Infection and Functions as an RNA Silencing Suppressor. *Journal of Virology*. 2016; 90:3543-3547.

Mondal S, Lin Y, Carroll J, Wenninger E, Bosque-Pérez N, Whitworth J *et al.* *Potato virus Y* Transmission Efficiency from Potato Infected with Single or Multiple Virus Strains. *Phytopathology*. 2017; 107:491-498.

Moury B. A new lineage sheds light on the evolutionary history of *Potato virus Y*. *Molecular Plant Pathology*. 2010; 11:161-168.

Moury B, Verdin E. Viruses of pepper crops in the Mediterranean basin: a remarkable stasis. *Advances in Virus Research*. 2012; 84:127–162.

Muñoz-Baena L, Gutiérrez-Sánchez P, Marín-Montoya M. Detección y secuenciación del genoma de *Potato virus Y* (PVY) que infecta plantas de tomate en Antioquia, Colombia. *Bioagro*. 2016; 28:69-80.

Neves-Borges A, Collares W, Pontes J, Breyne P, Farinelli L, De Oliveira D. Coat protein RNAs-mediated protection against *Andean potato mottle virus* in transgenic tobacco. *Plant Science*. 2001; 160:699–712.

Nie X, Singh R. A new approach for the simultaneous differentiation of biological and geographical strains of *Potato virus Y* by uniplex and multiplex RT-PCR. *Journal of Virological Methods*. 2002; 104:41-54.

Ode H, Nakashima M, Kitamura S, Sugiur W, Sato H. Molecular dynamics simulation in virus research. *Frontiers in microbiology*. 2012; 3:1-9.

Osorio I, Gutiérrez P, Marín M. *Spongospora subterranea* f.sp. *subterranea* y su Virus Asociado *Potato mop-top virus* (PMTV), dos Patógenos Reemergentes en los Cultivos de Papa de Colombia. *Revista Facultad Nacional de Agronomía*. 2012; 65:6361-6378.

Pacheco R, García-Marcos A, Manzano A, García de Lacoba M, Camañes G, García-Agustín P, *et al.* Comparative analysis of transcriptomic and hormonal responses to compatible and

incompatible plant-virus interactions that lead to cell death. *Molecular Plant Microbe Interaction*. 2012; 25:709-723.

Pasin F, Simón-Mateo C, García J. The hypervariable amino-terminus of P1 protease modulates potyviral replication and host defense responses. *PLoS Pathogens*. 2014; 10:1-16.

Perilla J, Hadden J, Goh B, Mayne C, Schulten K. All-Atom Molecular Dynamics of Virus Capsids as Drug Targets. *The Journal of Physical Chemistry Letters*. 2016; 7:1836-1844.

Petrov N, Stoyanova M, Andonova R, Teneva A. Induction of resistance to *Potato virus Y* strain NTN in potato plants through RNAi. *Biotechnology, biotechnological equipment*. 2015; 29:21-26.

Phan J, Zdanov A, Evdokimov A, Tropea J, Peters H, Kapust R, *et al*. Structural basis for the substrate specificity of tobacco etch virus protease. *Journal of Biological Chemistry*. 2002; 277:50564–50572.

Plisson C, Drucker M, Blanc S, German-Retana S, Le Gall O, Thomas D *et al*. Structural characterization of HC-Pro, a plant virus multifunctional protein. *Journal of Biological Chemistry*. 2003; 278:23753-23761.

Quenouille J, Vassilakos N, Moury, B. *Potato virus Y*: a major crop pathogen that has provided major insights into the evolution of viral pathogenicity. *Molecular Plant Pathology*. 2013; 14:439-452.

Reddy T, Sansom M. Computational virology: From the inside out. *Biochimica et biophysica acta*. 2016; 1858:1610-1618.

Revers F, García J. Molecular Biology of Potyviruses. *Advances in Virus Research*. 2015; 92:101-199.

Riechmann J, Cervera M, García JA. Processing of the *Plum pox virus* polyprotein at the P3-6K1 junction is not required for virus viability. *Journal of General Virology*. 1995; 76:951-956.

Rimbaud L, Dallot S, Gottwald T, Decroocq V, Jacquot E, Soubeyrand S, *et al.* Sharka epidemiology and worldwide management strategies: learning lessons to optimize disease control in perennial plants. *Annual review of Phytopathology*. 2015; 53:357-378.

Rodríguez P, Franco L, Guzmán M. Inter and Intra Variation of *Potato Yellow Vein Virus* in Three Potato Species From Colombia. *Revista Facultad Nacional de Agronomía*. 2015; 68:7387–7398.

Rolland M, Lacroix C, Blanchard A, Baldwin T, Kerlan C, Jacquot E. *Potato virus Y (PVY): From its discovery to the latest outbreaks*. *Virologie*. 2008; 2:261-273.

Roossinck M. Plants, viruses and the environment: Ecology and mutualism. *Virology*. 2015; 1:479-480.

Rybicki E, Shukla D. Coat protein phylogeny and systematics of potyviruses. En: O. Barnett, Jr., editor. *Potyvirus taxonomy*: Viena: Springer; 1992. p. 139-170.

Salazar L, Mink G, Lecoq H, Lot H, Kegler H, Martelli G, *et al.* En: Milne R, editor. *The Plant Viruses*. Alemania: Springer; 1988:331-350.

Salazar L. *Potato Viruses after The XXth Century: Effects, Dissemination and Their Control*. Crop Protection Department. 2003.

Sánchez de Luque C, Corzo P, Pérez O. Incidencia de virus en papa y su efecto sobre rendimientos en tres zonas agroecológicas de Colombia. *Revista Latinoamericana de la Papa*. 1991; 4:36-51.

Sasaya T, Torrance L, Cowan G, Ziegler A. Aphid transmission studies using helper component proteins of *Potato virus Y* expressed from a vector derived from *Potato virus X*. *Journal of General Virology*. 2000; 81:1115–1119.

Scholthof K, Adkins S, Czosnek H, Palukaitis P, Jacquot E, Hohn T, *et al.* Top 10 plant viruses in molecular plant pathology. *Molecular Plant Pathology*. 2011; 12:938-954.

Schramm S, Frost K, Charkowski A, Gray S, Crockford A, Groves R. Management of *Potato virus Y* (PVY) in Wisconsin Seed Potato Production. University of Wisconsin-Extension. 2011. Disponible en: <https://learningstore.uwex.edu/Assets/pdfs/A3951.pdf> (Acceso noviembre 2016).

Sepúlveda P, Rojas C, Rosales M, Sepúlveda G. Manejo de enfermedades virales transmitidas por insectos en tomates en la región de Arica y Parinacota. *Boletín INIA* 224. 2009. Disponible en: <http://www2.inia.cl/medios/biblioteca/boletines/NR38129.pdf> (Acceso mayo 2016).

Shukla D, Ward C, Brunt A. The *Potyviridae*. *New Phytologist*. 1995; 2:289-290.

Singh M, Singh R, Fageria M, Nie X, Coffin R, Hawkins G. Optimization of a Real-Time RT-PCR assay and its comparison with ELISA, conventional RT-PCR and the grow-out test for large scale diagnosis of *Potato virus Y* in dormant potato tubers. *American Journal of Potato Research*. 2013; 90:43-50.

Singh R, Valkonen J, Gray S, Boonham N, Jones R, Kerlan C, Schubert J. The naming of *Potato virus Y* strains infecting potato. *Archives of Virology*. 2008; 153:1-13.

Smith K. On the Composite Nature of Certain Potato Virus Diseases of the Mosaic Group as Revealed by the use of Plant Indicators and Selective Methods of Transmission. . 1931; 109:251-266.

Sorel M, Garcia J, German-Retana S. The *Potyviridae* Cylindrical Inclusion Helicase: A Key Multipartner and Multifunctional Protein. *The American Phytopathological Society*. 2014; 27:215-226.

Šubr Z, Kamencayová M, Nováková S, Nagyová A, Nosek J, Glasa M. A single amino acid mutation alters the capsid protein electrophoretic double-band phenotype of the *Plum pox virus* strain PPV-Rec. *Archives of Virology*. 2010; 155:1151-1155.

Syller J. The roles and mechanisms of helper component proteins encoded by potyviruses and caulimoviruses. *Physiological and Molecular Plant Pathology*. 2006; 67:119–130.

Tennant P, Fermin G, Roye M. Viruses infecting papaya (*Carica papaya* L.) etiology, pathogenesis and molecular biology. *Plant Viruses*. 2007; 1:178-188.

Toro E, Santa J, Granada M. Solution of the vehicle routing problem for the potato distribution in Colombia. *Scientia et Technica Año XVIII*. 2013; 18:139–149.

Ulf H. Thermodynamic driving forces in protein regulation studied by molecular dynamics simulations (tesis de doctorado). Gotinga: Facultad de Matemáticas y Ciencias Naturales, Universidad de Gotinga; 2008:9-15.

Unimedios. Tratamiento con microorganismos para erradicar la sarna polvosa de la papa. 2008. Disponible en: <http://historico.agenciadenoticias.unal.edu.co> (Acceso mayo 2016).

USDA. Managing *Potato virus Y* in Seed Potato Production. Ithaca, NY. 2016. Disponible en <http://www.potatovirus.com/index.cfm/page/PVYinfo.htm> (Acceso mayo 2016).

Velásquez N, Hossaing M, Murphy J. Differential disease symptoms and full-length genome sequence analysis for three strains of *Tobacco etch virus*. *Virus genes*. 2015; 50:442-449.

Valkonen J. Elucidation of virus-host interactions to enhance resistance breeding for control of virus diseases in potato. *Breeding science*. 2015; 65:69-76.

Valkonen J, Slack S, Plaisted R, Atanabe K. Extreme resistance is epistatic to hypersensitive resistance to *Potato virus Y<sup>O</sup>* in a *Solanum tuberosum* subsp. andigena-derived potato genotype. *Plant Disease*. 1994; 78:1177–1180.

Valli A, Gallo A, Calvo M, Pérez Jd.J, García, J. A novel role of the potyviral helper component proteinase contributes to enhance the yield of viral particles. *Journal of Virology*. 2014; 88:9808-9818.

Valli A, López-Moya J, García J. Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family *Potyviridae*. *Journal of General Virology*. 2007; 88:1016-1028.

Verbeek M, Piron P, Dullemans A, Cuperus C, van derVlugt R. Determination of aphid transmission efficiencies for N, NTN and Wilga strains of *Potato virus Y*. *Annals of Applied Biology*. 2010; 156:39–49.

Villamil-Garzón A, Cuellas W, Guzmán-Barney M. Co-infección natural de *Potato yellow vein virus* y potivirus en cultivos de *Solanum tuberosum* en Colombia. *Agronomía Colombiana*. 2014; 32:213-223.

Visser J, Bellstedt D, Pirie M. The recent recombinant evolution of a major crop pathogen, *Potato virus Y*. *PLoS One*. 2012; 7:1-12.

Waltermann A, Maiss E. Detection of 6K1 as a mature protein of 6 kDa in plum pox virus-infected *Nicotiana benthamiana*. *Journal of General Virology*. 2006; 87:2381-2386.

White K. The polymerase slips and PIPO exists. *EMPO Report*. 2015; 16:885-886.

Yoon H, Hwang D, Choi K, Song B. Proteolytic Processing of Oligopeptides containing the target sequences by the Recombinant *Tobacco vein mottling virus* Nia proteinase. *Molecular Cells*. 1999; 10:213-219.

Yoshida N, Shimura H, Yamashita K, Suzuki M, Masuta C. Variability in the P1 gene helps to refine phylogenetic relationships among *Leek yellow stripe virus* isolates from garlic. *Archives of Virology*. 2012; 157:147-153.

Yotmanee P, Rungrotmongkol T, Wichapong K, Choi S, Wahab H, Kungwan N, *et al.* Binding specificity of polypeptide substrates in NS2B/NS3pro serine protease of dengue virus type 2: A molecular dynamics Study. *Journal of molecular graphics and modelling.* 2015; 60:24-33.

### 3. OBJETIVOS

#### 3.1. General

Caracterizar los genomas de las principales variantes de PVY que afectan los cultivos de *S. tuberosum* en Antioquia utilizando métodos de secuenciación de nueva generación y Dinámica Molecular.

#### 3.2. Específicos

- Evaluar los niveles de incidencia y diversidad genética de PVY en cultivos de *S. tuberosum* del Oriente y Norte de Antioquia, utilizando métodos convencionales de biología molecular y NGS.
- Obtener secuencias completas de los genomas de variantes de PVY infectando *S. tuberosum* en Antioquia, mediante métodos de secuenciación masiva de nueva generación.
- Caracterizar el mecanismo de reconocimiento de la proteasa NIa-Pro en sus siete sitios de corte: P3/6K1, 6K1/CI, CI/6K2, 6K2/VPg, VPg/NIa-Pro, NIa-Pro/NIb, NIb/CP, utilizando métodos de Dinámica Molecular.
- Obtener un modelo cinético del procesamiento proteolítico de potyvirus mediado por NIa-Pro utilizando la información cinética derivada de simulaciones por Dinámica Molecular.

## 4. Resultados

### 4.1. Capítulo 1. Análisis del mecanismo de reconocimiento de Nla-Pro para la postulación de un modelo del procesamiento postraduccional de la poliproteína en TEV y PVY, utilizando simulaciones por Dinámica Molecular.

#### 4.1.1. Introducción

El procesamiento de la poliproteína en los potyvirus es controlado a través de variaciones en las secuencias consenso que permiten una tasa de producción individual para intermediarios y proteínas finales. Con el propósito de caracterizar el mecanismo de reconocimiento de la proteasa Nla-Pro en sus siete sitios de corte, se construyeron modelos por homología para las proteasas de TEV y PVY, evaluando el comportamiento de los sitios de corte para cada uno de los virus y empleando simulaciones por Dinámica Molecular. Finalmente, los resultados obtenidos permitieron establecer un modelo de maduración cinética utilizando estimaciones de las tasas de corte a partir de las simulaciones. La validez de las predicciones fue corroborada con resultados obtenidos en experimentos de proteasas mutantes de potyvirus reportadas en la literatura.

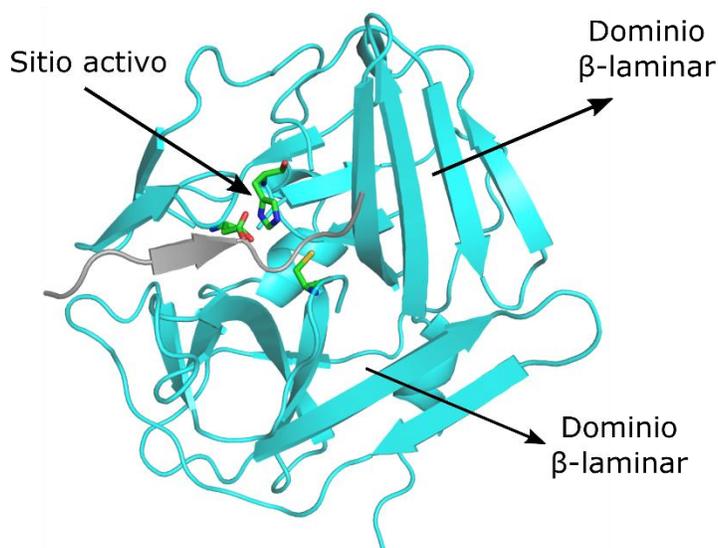
#### 4.1.2. Descripción del mecanismo de corte: paralelo entre Nla-Pro y tripsina (serín-proteasa)

Las cisteín-proteasas se encuentran dentro de las siete mayores clases de enzimas proteolíticas. Conforman un grupo de 21 familias en las que se destacan: la clostripaína (*Clostridium histolyticum*), catepsina (*Aspergillus flavus*), cruzipaína (*Trypanosoma cruzi*) y la papaína, fuertemente caracterizada en plantas (*Carica papaya*) (Grzonka *et al.*, 2001). Sin embargo, por lo menos la mitad de todas las cisteín-proteasas han sido descubiertas en virus; entre ellas la Nla-Pro de *Tobacco etch virus* (TEV) se destaca por su rigurosa especificidad de secuencia en el corte y por ser la más estudiada entre los potyvirus debido a su alta actividad y especificidad. Esta enzima es utilizada ampliamente en aplicaciones biotecnológicas como el procesamiento de polipéptidos y en la remoción de marcapos para

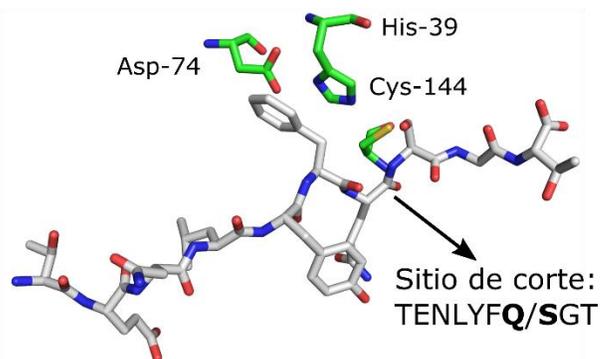
purificación por afinidad de proteínas recombinantes (Cesaratto *et al.*, 2016).

Se sabe que en los potyvirus el genoma de RNA codifica por lo menos para dos marcos de lectura (ORF) que son procesados por las proteasas virales P1, HC-Pro y NIa-Pro. El corte de la poliproteína de mayor tamaño resulta en diez proteínas maduras: P1, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb y CP. Por su parte, la poliproteína de menor tamaño genera solamente tres proteínas maduras: P1, HC-Pro y P3N-PIPO. En el 2002, se resolvieron las estructuras cristalográficas de dos proteasas mutantes para TEV: una catalíticamente inactiva (C151A) en asociación con el sustrato y otra resistente a la autólisis (S219D) en complejo con el producto, revelando las bases estructurales de su actividad (Phan *et al.*, 2002). Como se observa en la Figura 11-A, NIa-Pro adopta un plegamiento de dos dominios  $\beta$ -laminares antiparalelos en cuya interfaz se encuentran los residuos His-39, Asp-74 y Cys-144 que constituyen su sitio activo (Figura 11-B). Esta conformación explica la reactividad de la cisteína 144, pues su grupo tiol forma un puente de hidrógeno con el nitrógeno  $\epsilon_2$  del anillo imidazol de la histidina 39. El grupo amino en la posición  $\delta_1$  de la histidina forma, a su vez, un puente de hidrógeno con el grupo carboxilo del ácido aspártico 74, constituyendo la tríada catalítica (Figura 12).

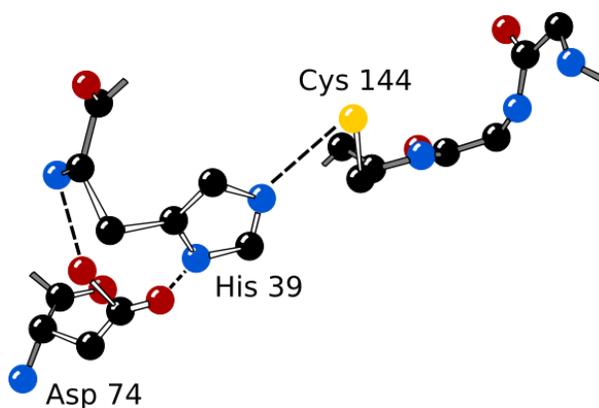
**A**



**B**



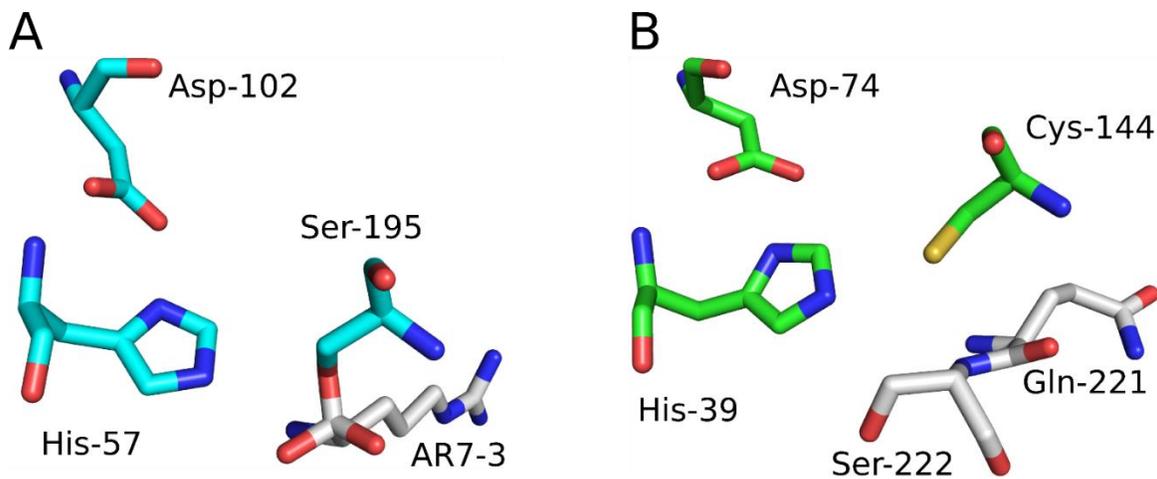
**Figura 11. Proteasa NIa-Pro catalíticamente activa de *Tobacco etch virus*. A)** Estructura cristalográfica de la NIa-Pro en complejo con el sustrato (gris). Los residuos coloreados por átomos (verde) conforman la tríada catalítica que se encuentra entre los dos dominios  $\beta$ -laminares. **B)** Secuencia del péptido-sustrato NIb/CP coloreado por átomos (blanco) en presencia de la tríada catalítica (verde).



**Figura 12. Residuos de la tríada catalítica de Nla-Pro de TEV.** Los residuos involucrados en la actividad catalítica de la proteasa Nla-Pro se distinguen mediante enlaces en color blanco, de los enlaces en color gris del resto de la proteína.

El residuo de histidina se encuentra cerca a la cadena lateral de la cisteína 144, dejando preparado a su grupo tiol para una desprotonación. Cuando el sustrato está presente, la histidina acepta dicho protón lo que deja al azufre con una fuerte propiedad nucleofílica. La tarea del ácido aspártico en tal evento consiste en facilitar la protonación de la histidina y estabilizar la carga positiva generada mediante efectos electrostáticos.

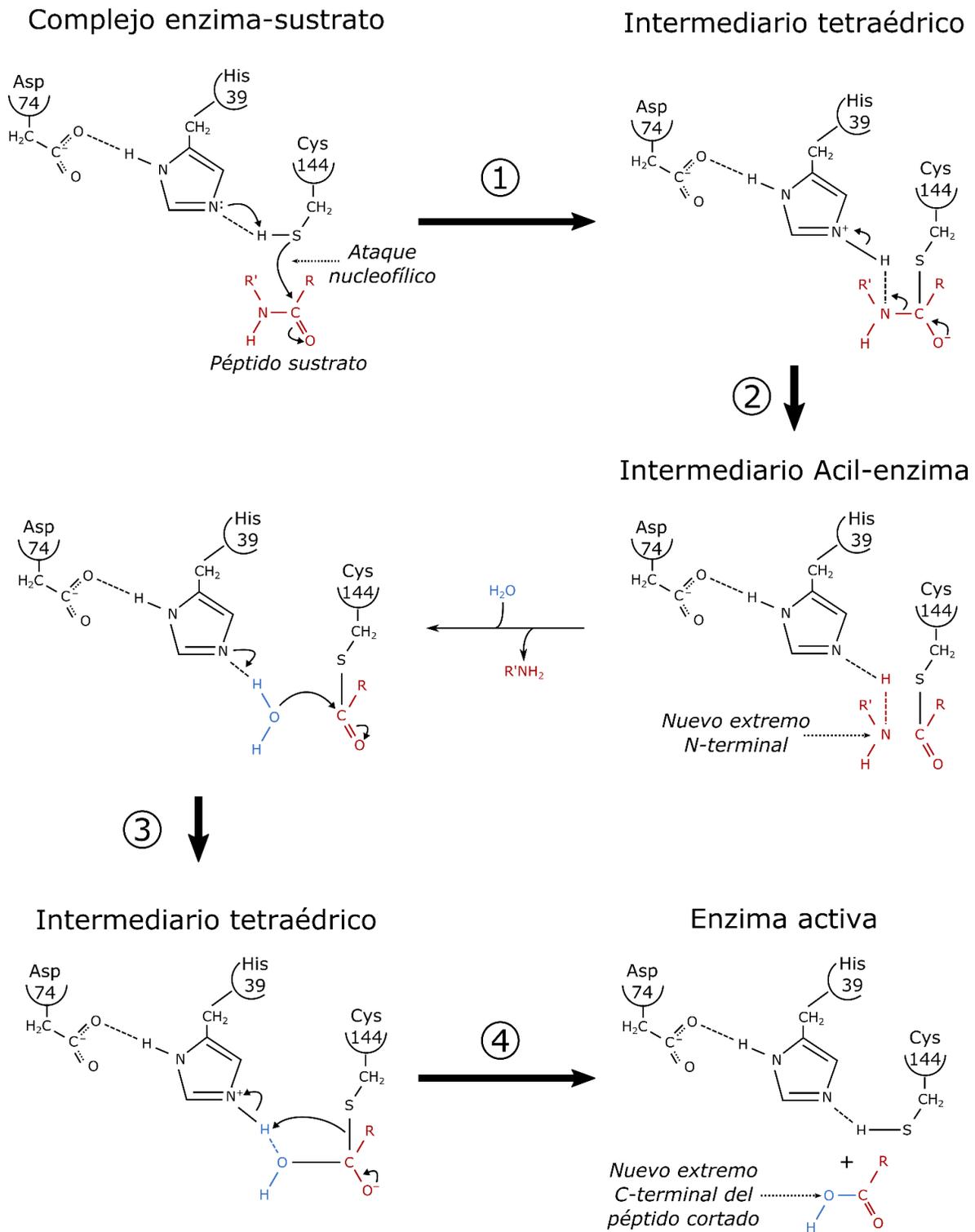
El sitio activo de Nla-Pro presenta una gran similitud con las serín-proteasas. Este grupo de enzimas se caracteriza por la presencia de la serina, en lugar de cisteína, en el sitio activo (Berg *et al.*, 2007). La reacción de esta enzima puede resumirse en los siguientes pasos: en primer lugar (acilación), ocurre un ataque nucleofílico desde la serina al enlace peptídico propenso al corte, formando un intermediario tetraédrico y una acil-enzima que libera el extremo C-terminal del péptido. Posteriormente, tras una reacción de deacilación una molécula de agua ataca la acil-enzima generando un segundo intermediario tetraédrico seguido de la liberación el extremo N-terminal (Radisky *et al.*, 2006). En la Figura 13-A se ilustra el entorno molecular de la tríada catalítica de la tripsina (serín-proteasa) conformada por los residuos His-57, Asp-102 y Ser-195, en complejo con la leupeptina; adicionalmente se realiza un paralelo con la Nla-Pro de TEV (Figura 13-B).



**Figura 13. Residuos de la tríada catalítica de las serín y cisteín-proteasas. A)** Tripsina en complejo con la leupeptina como péptido sustrato coloreado por átomos (blanco). Los residuos de la tríada catalítica se colorean por átomos en cian. En este caso se representa el intermediario tetraédrico gracias al complejo entre la tripsina, la leupeptina y un aldehído-oligopéptido que funciona como inhibidor, AR7-3 (PDB ID: 2AGI). **B)** Nla-Pro de TEV en complejo con residuos del sitio de corte N1b/CP (blanco). Los aminoácidos de la tríada catalítica se colorean por átomos en verde.

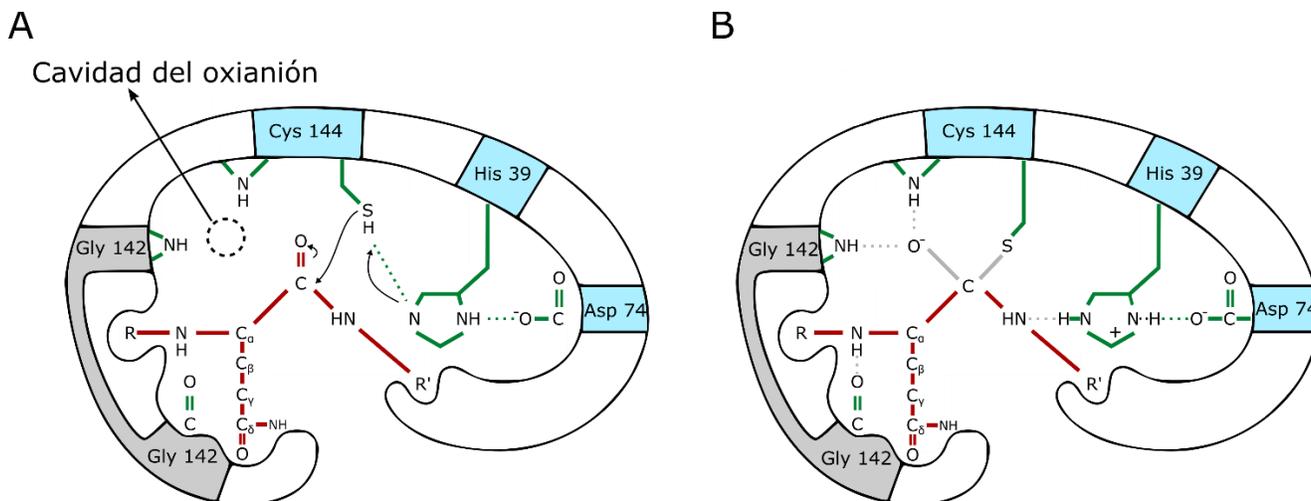
La similitud entre los sitios activos de estas dos enzimas sugiere mecanismos catalíticos análogos. En las cisteín-proteasas, luego de la unión del sustrato, ocurre un ataque nucleofílico por parte del azufre hacia el carbono carbonílico del enlace peptídico. Posteriormente, la unión de cuatro átomos al carbono promueve el paso de una disposición planar a un intermediario tetraédrico, cuya inestabilidad mantiene una carga negativa sobre el átomo de oxígeno (Figura 14, paso 1). La estabilización sobre este grupo se lleva a cabo mediante interacciones con grupos NH de la proteína (glicina 143) en un centro denominado cavidad del oxianión (Figura 15) (Berg *et al.*, 2007). Estas interacciones también intervienen en la estabilidad del estado de transición que precede al intermediario tetraédrico (Figura 16). La ruptura del intermediario tetraédrico resulta en la generación de un intermediario acil-enzima gracias a la transferencia de un protón desde la histidina cargada positivamente al grupo amino, formado por la ruptura del enlace peptídico, que ahora se encuentra libre para apartarse de la enzima. En el paso siguiente, sucede una reacción de deacilación en la que una molécula de agua se sitúa en el lugar previamente ocupado por el grupo amino del sustrato. El grupo éster de la acil-enzima se hidroliza mediante un proceso igual que en el paso 1. El ion hidróxido, resultante de la toma del protón por parte de la histidina 39, ataca el átomo carbonílico del grupo acilo formando un segundo intermediario tetraédrico (Figura 14,

paso 3). La descomposición de esta estructura forma el producto ácido carboxílico, cuya liberación permite desarrollar otro ciclo catalítico de la enzima (Figura 14, paso 4).

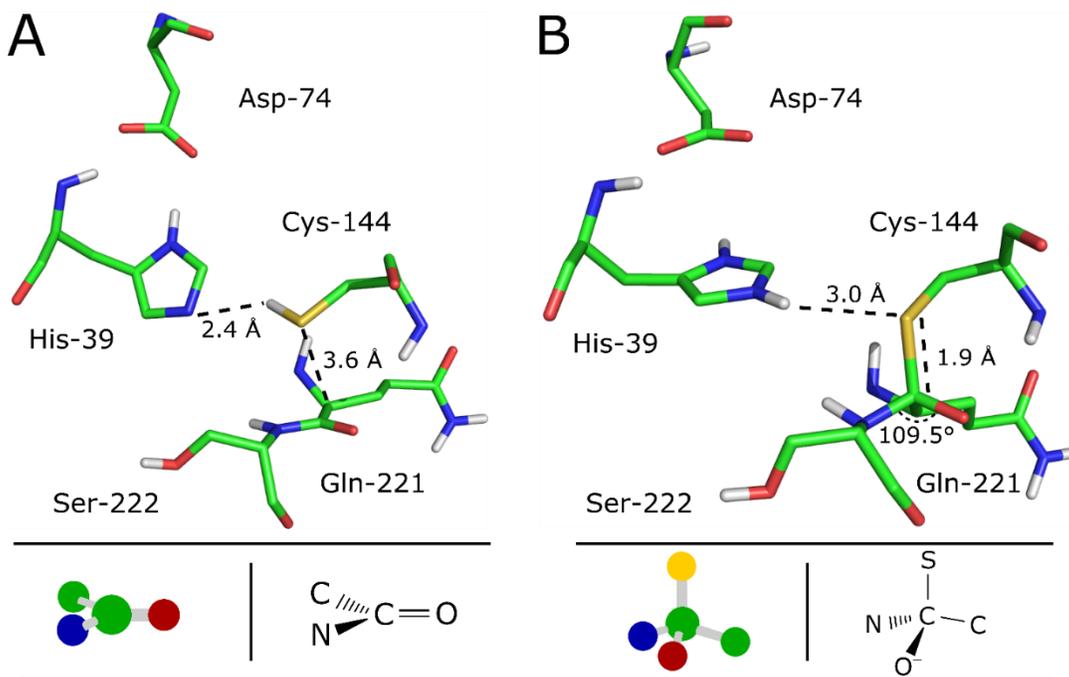


**Figura 14. Mecanismo catalítico postulado para la N1a-Pro de TEV.** La reacción involucra (1) un ataque nucleofílico desde el sitio activo de la cisteína 144 hacia el grupo carbonilo del enlace peptídico del sustrato propenso a ser cortado, para formar un primer intermediario tetraédrico. (2)

Descomposición del intermediario en un complejo acil-enzima mediado por la catálisis de la histidina 39 (una vez polarizada por la acción del ácido aspártico 74), seguido de la pérdida del grupo amino y la entrada de una molécula de agua que lo reemplaza. **(3)** En la reacción de deacilación se forma un segundo intermediario tetraédrico. **(4)** Resultado del producto carboxilo y regeneración de la enzima activa.



**Figura 15. Estabilización del estado de transición en la Nla-Pro de TEV. A)** En el complejo Michaelis, el carbono carbonílico trigonal del enlace peptídico presenta una restricción para su unión a la cavidad del oxianión (parte central superior con enlaces en rojo). **B)** En el intermediario tetraédrico, la carga negativa del oxígeno del grupo carbonilo (*scissile bond*) ahora puede formar puentes de hidrógeno con los grupos NH de los residuos glicina-142 y cisteína-144 de la proteína (enlaces en color gris). Los residuos de la tríada catalítica se resaltan en color azul.



**Figura 16. Geometría del carbono carbonílico en el enlace peptídico.** En **A)** se aprecia la disposición planar del enlace peptídico, cuyo carbono presenta una hibridización  $sp^2$ . **B)** Durante el intermediario tetraédrico que sucede a la desprotonación del grupo tiol, el carbono carbonílico del enlace peptídico adopta una conformación tetragonal  $sp^3$  una vez ocurre el ataque nucleofílico.

### 4.1.3. Modelación por homología

La estructura cristalográfica de Nla-Pro de TEV se utilizó como molde para la construcción de una proteasa homóloga funcional, junto con el respectivo péptido-sustrato (sitio de corte) y la mutación que restaura su actividad catalítica (A151C). La Figura 17 muestra el formato de alineamiento empleado para una molécula de dos cadenas en el sitio de corte VPg/Nla-Pro. Allí, se especifican los nombres del archivo de entrada (1LVBs), el archivo de salida (por ejemplo: 7\_TEV) y la región de la proteasa a modelar. La secuencia del péptido es señalada por la separación de la cadena principal y está conformada por diez residuos: siete antes del corte y tres después (Merits *et al.*, 2002; Yoon *et al.*, 1992). Posteriormente, se utilizó un script de MODELLER para obtener un total de diez modelos de la proteasa Nla-Pro de TEV catalíticamente activa.

**A**

```
>P1;1LVBs
structureX:1LVBs: FIRST :A:LAST :C:::
-----
PRDYNPISSTICHLTNE SDGHTTSLYGIGFGPFIITNKHLFRRNNGT
LLVQSLHG VFKVKN TTTLQQHLIDGRD MIIIRMPKDFPPFPQKLF
REPQREERICLVTTNFQTKSMSSMVSDT SCTFPSSDGIFWKHWIQ
TKDGGQAGSPLVSTRDGFIVGIHSASNFTNTNNYFTSVPKNFM-ELL
TNQEAQQWVSGWRLNADSVLWGGHKVFMSKP-----
-----/TENLYFQSGT*

>P1;7_TEV
sequence:7_TEV:
FIRST :A:LAST:C:undefined:undefined:-1.00:-1.00
-----
PRDYNPISSTICHLTNE SDGHTTSLYGIGFGPFIITNKHLFRRNNGT
LLVQSLHG VFKVKN TTTLQQHLIDGRD MIIIRMPKDFPPFPQKLF
REPQREERICLVTTNFQTKSMSSMVSDT SCTFPSSDGIFWKHWIQ
TKDGGQCGSPLVSTRDGFIVGIHSASNFTNTNNYFTSVPKNFM-ELL
TNQEAQQWVSGWRLNADSVLWGGHKVFMSKP-----
-----/NEDLTFEGES*
```

**B**

```
# Comparative modeling by the automodel class

from modeller import *
from modeller.automodel import *

log.verbose()
env = environ()

env.io.atom_files_directory = ['.', '..']

a = automodel(env,
              alnfile = 'aln.fas',           ← Archivo de alineamiento
              knowns = '1LVBs',            ← Código del modelo de referencia
              sequence = '7_TEV')          ← Nombre estructura de salida

a.starting_model= 1
a.ending_model = 10                        ← Número de estructuras creadas

a.make()
```

**Figura 17. Archivos requeridos por MODELLER para la modelación por homología.** **A)** Archivo de alineamiento para la construcción de un modelo por homología de dos cadenas. La alanina en el panel superior es sustituida por cisteína en el panel inferior (ambos en negrilla). El péptido que incluye el sitio de corte (azul) se ubica después del separador de cadenas (/) y finaliza en \*. **B)** Script *model-multichain-sym.py* modificado para obtener 10 modelos por homología catalíticamente activos según la estructura de la proteasa de TEV (PDB IB: 1LVB); los archivos requeridos para la construcción del modelo se señalan con flechas en la imagen.

Esta metodología se repitió para cada uno de los sitios de corte de los cuatro genomas virales, según las secuencias reportadas en la Figura 18. Como se indica en la Figura 19, la salida del programa permite definir la mejor estructura de acuerdo al menor valor de la propiedad *molecular PDF* (molpdf), que corresponde a una función de puntajes interna de MODELLER.



Sitio de corte	TEV	f_PVY	t_PVY
P1/HC-Pro	CHSMTHY/SDK	MDSMVQF/GVM	LDSMVQF/SAE
HC-Pro/P3	MKTYNVG/GMN	IKHYRVG/GIP	IKHYRVG/GIP
P3/6K1	KEDLVEQ/AKQ	EYDVRHQ/RST	EYDVRHQ/RST
6K1/CI	REIIYTQ/SLD	DHEVRHQ/SLD	DHEVRHQ/SLD
CI/6K2	LETIYLQ/SDS	LQFVHHQ/AAT	LQFVHHQ/ATT
6K2/VPg	NEPVYFQ/GKK	VETVSHQ/GKN	VETVSHQ/GRN
VPg/NIa-Pro	NEDLTFE/GES	KQEVEHE/AKS	KQEVEHE/AKS
NIa-Pro/Nib	NELVYSQ/GEK	HDVVVEQ/AKH	HDVVVEQ/AKH
Nib/CP	TENLYFQ/SGT	TYEVHHQ/GND	TYEVHHQ/GND

**Figura 18. Secuencias de las regiones de corte para TEV y PVY.** En la parte superior se ubica la estructura de un genoma de potyvirus, indicando los sitios de corte realizados por NIa-Pro mediante flechas. Las secuencias correspondientes a estos sitios se muestran en la tabulación inferior. f\_PVY: PVY secuenciado a partir de tejido foliar; t\_tVY: PVY secuenciado a partir de tubérculo.

```
>> Summary of successfully produced models:
Filename                               molpdf
-----
7_TEV.B99990001.pdb                   1395.88477
7_TEV.B99990002.pdb                   1335.91125
7_TEV.B99990003.pdb                   1460.58704
7_TEV.B99990004.pdb                   1296.83936
7_TEV.B99990005.pdb                   1417.77246
7_TEV.B99990006.pdb                   1432.91321
7_TEV.B99990007.pdb                   1379.94067
7_TEV.B99990008.pdb                   1291.42615
7_TEV.B99990009.pdb                   1298.03845
7_TEV.B99990010.pdb                   1435.76538
```

**Figura 19. Archivo de reporte de MODELLER.** La construcción de los modelos por homología finaliza con un reporte de energías para cada una de las estructuras construidas. La mejor de ellas se determinó según el menor valor de *molpdf*, en este caso correspondió al octavo resultado (señalado en negrita).

#### 4.1.4. Simulaciones por Dinámica Molecular (MD)

Las simulaciones por Dinámica Molecular en NAMD requieren de un archivo en formato PDB (*Protein Data Bank*) con las coordenadas de cada uno de los átomos en la estructura (Figura 20) y un archivo PSF (*Protein Structure File*) que contiene información sobre los átomos (*atoms*), enlaces (*bonds*), ángulos (*angles*), ángulos diedros (*dihedrals*), torsiones (*impropers*) y términos cruzados (*cross-terms*) necesarios para la aplicación del campo de fuerzas durante la simulación. Adicionalmente, se necesita de un archivo de topología (*top\_all27\_prot\_lipid.inp*) y otro de parámetros (*par\_all27\_prot\_lipid.inp*). El archivo de topología contiene la información necesaria para convertir nombres de residuos en archivos completos de estructuras, mediante el tipo, la masa y la carga de cada átomo en cada residuo; también almacena coordenadas internas que permiten la asignación automática de los hidrógenos y otros átomos faltantes en el archivo PDB. Las moléculas de agua e iones de sodio o cloro también se encuentran especificados en el archivo de topología. Finalmente, el archivo de parámetros contiene las constantes numéricas necesarias para evaluar fuerzas y energías con base en los archivos de estructuras y de coordenadas. En la Figura 21 se ilustra un segmento del archivo de topología para la prolina.

Archivo PDB										Archivo PSF																
ATOM	1	N	PRO	A	1	57.277	41.054	36.028	1.00	61.10	N	16709 !NATOM														
ATOM	2	CA	PRO	A	1	56.933	42.190	35.132	1.00	61.10	C	1	P1	1	PRO	NN1	HC	-0.070000	14.0070	0						
ATOM	3	CD	PRO	A	1	57.739	39.887	35.202	1.00	61.10	C	2	P1	1	PRO	NN2	HC	0.240000	1.0080	0						
ATOM	4	CB	PRO	A	1	56.987	41.643	33.708	1.00	61.10	C	3	P1	1	PRO	HD1	HA	0.090000	1.0080	0						
ATOM	5	CG	PRO	A	1	57.935	40.440	33.784	1.00	61.10	C	4	P1	1	PRO	CD	CP3	0.160000	12.0110	0						
ATOM	6	C	PRO	A	1	55.574	42.634	35.542	1.00	61.10	C	5	P1	1	PRO	HD1	HA	0.090000	1.0080	0						
ATOM	7	O	PRO	A	1	54.750	41.794	35.901	1.00	61.10	O	6	P1	1	PRO	HD2	HA	0.090000	1.0080	0						
ATOM	8	N	ARG	A	2	55.325	43.953	35.508	1.00104.55	N	7	P1	1	PRO	CA	CP1	0.160000	12.0110	0							
ATOM	9	CA	ARG	A	2	54.055	44.470	35.912	1.00104.55	C	8	P1	1	PRO	HA	HB	0.090000	1.0080	0							
ATOM	10	CB	ARG	A	2	54.143	45.958	36.289	1.00104.55	C	9	P1	1	PRO	CB	CP2	-0.180000	12.0110	0							
ATOM	11	CG	ARG	A	2	53.115	46.412	37.323	1.00104.55	C	10	P1	1	PRO	HB1	HA	0.090000	1.0080	0							
ATOM	12	CD	ARG	A	2	51.706	46.648	36.792	1.00104.55	C	11	P1	1	PRO	HB2	HA	0.090000	1.0080	0							
ATOM	13	NE	ARG	A	2	50.892	47.073	37.965	1.00104.55	N	12	P1	1	PRO	CG	CP2	-0.180000	12.0110	0							
ATOM	14	CZ	ARG	A	2	49.574	47.383	37.818	1.00104.55	C	13	P1	1	PRO	HG1	HA	0.090000	1.0080	0							
ATOM	15	NH1	ARG	A	2	48.986	47.299	36.588	1.00104.55	N	14	P1	1	PRO	HG2	HA	0.090000	1.0080	0							
ATOM	16	NH2	ARG	A	2	48.839	47.781	38.898	1.00104.55	N	15	P1	1	PRO	C	C	0.510000	12.0110	0							
ATOM	17	C	ARG	A	2	53.156	44.343	34.721	1.00104.55	C	16	P1	1	PRO	O	O	-0.510000	15.9990	0							
ATOM	18	O	ARG	A	2	53.601	44.528	33.590	1.00104.55	O																
ATOM	19	N	ASP	A	3	51.866	44.012	34.937	1.00	49.54	N	12351 !NBOND: bonds														
ATOM	20	CA	ASP	A	3	50.946	43.867	33.844	1.00	49.54	C	1		7		1		4		2		1		3		1
ATOM	21	CB	ASP	A	3	49.939	42.721	34.058	1.00	49.54	C	5		4		6		4		7		9		8		7
ATOM	22	CG	ASP	A	3	49.140	42.527	32.778	1.00	49.54	C	9		12		10		9		11		9		12		4
ATOM	23	OD1	ASP	A	3	49.332	43.333	31.829	1.00	49.54	O	13		12		14		12		15		7		15		17
ATOM	24	OD2	ASP	A	3	48.323	41.569	32.733	1.00	49.54	O	16		15		17		18		17		19		19		20
ATOM	25	C	ASP	A	3	50.158	45.139	33.732	1.00	49.54	C	21		19		21		22		21		23		24		21
ATOM	26	O	ASP	A	3	49.305	45.435	34.566	1.00	49.54	O	24		25		24		26		27		24		27		28
ATOM	27	N	TYR	A	4	50.495	45.941	32.706	1.00105.96	N	27		29		30		27		30		31		32		30	
ATOM	28	CA	TYR	A	4	49.936	47.217	32.349	1.00105.96	C	32		33		33		34		33		35		36		32	
ATOM	29	CB	TYR	A	4	50.898	48.046	31.486	1.00105.96	C	36		37		36		38		39		19		39		41	

**Figura 20. Formato de los archivos PDB y PSF utilizados en las simulaciones moleculares.** El archivo PDB muestra el primer aminoácido según la estructura secundaria de la proteasa Nla-Pro modelada para TEV. En este caso, la información sobre la prolina se resume en doce columnas donde las coordenadas x, y, z de cada átomo en dicho residuo, corresponden a las columnas siete, ocho y nueve, respectivamente. El archivo PSF fue creado con el paquete *psfgen* (extensión *Automatic PSF Builder*) del software de visualización molecular VMD. En el panel superior e inferior se muestran las secciones de *átomos* y *enlaces*, respectivamente. Los hidrógenos creados se señalan en gris, los cuáles no figuran en el archivo PDB.

```

RESI PRO          0.00
GROUP            !
ATOM N           N   -0.29 ! | \ /
ATOM CD          CP3  0.00 ! | N---CD HG1
ATOM HD1         HA   0.09 ! | \ /
ATOM HD2         HA   0.09 ! | \ /
ATOM CA          CP1  0.02 ! | \ /
ATOM HA          HB   0.09 ! | HA-CA--CB HG2
GROUP            !
ATOM CB          CP2 -0.18 ! | HB1 HB2
ATOM HB1         HA   0.09 ! O=C
ATOM HB2         HA   0.09 ! |
GROUP
ATOM CG          CP2 -0.18
ATOM HG1         HA   0.09
ATOM HG2         HA   0.09
GROUP
ATOM C           C    0.51
ATOM O           O   -0.51
BOND C CA C +N
BOND N CA CA CB CB CG CG CD N CD
BOND HA CA HG1 CG HG2 CG HD1 CD HD2 CD HB1 CB HB2 CB
DOUBLE O C
IMPR N -C CA CD
IMPR C CA +N O
CMAP -C N CA C N CA C +N
ACCEPTOR O C
IC -C CA *N CD 1.3366 122.9400 178.5100 112.7500 1.4624
IC -C N CA C 1.3366 122.9400 -76.1200 110.8600 1.5399
IC N CA C +N 1.4585 110.8600 180.0000 114.7500 1.3569
IC +N CA *C O 1.3569 114.7500 177.1500 120.4600 1.2316
IC CA C +N +CA 1.5399 116.1200 180.0000 124.8900 1.4517
IC N C *CA CB 1.4585 110.8600 113.7400 111.7400 1.5399
IC N C *CA HA 1.4585 110.8600 -122.4000 109.0900 1.0837
IC N CA CB CG 1.4585 102.5600 31.6100 104.3900 1.5322
IC CA CB CG CD 1.5399 104.3900 -34.5900 103.2100 1.5317
IC N CA CB HB1 1.4585 102.5600 -84.9400 109.0200 1.1131
IC N CA CB HB2 1.4585 102.5600 153.9300 112.7400 1.1088
IC CA CB CG HG1 1.5399 104.3900 -156.7200 112.9500 1.1077
IC CA CB CG HG2 1.5399 104.3900 81.2600 109.2200 1.1143
IC CB CG CD HD1 1.5322 103.2100 -93.5500 110.0300 1.1137
IC CB CG CD HD2 1.5322 103.2100 144.5200 110.0000 1.1144
PATCHING FIRS PROP

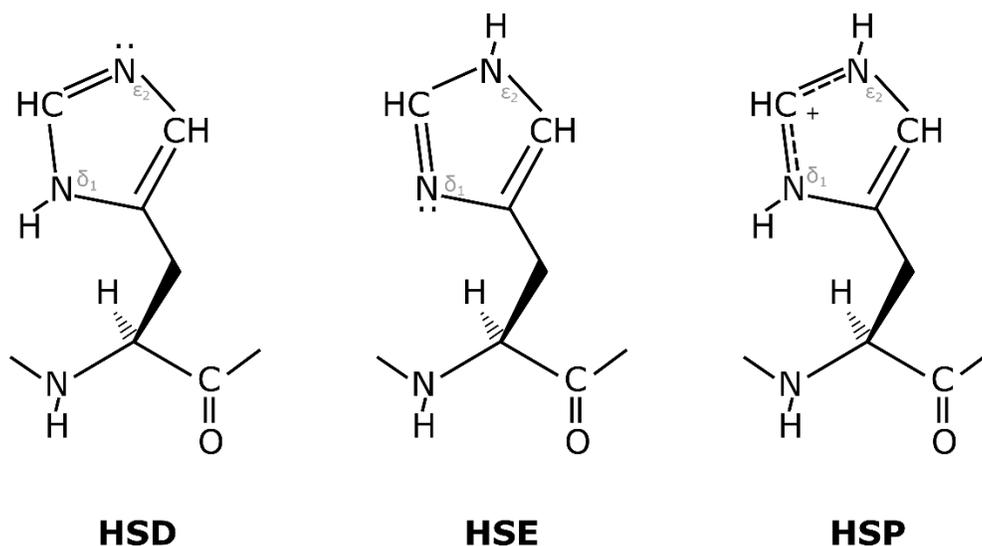
```

**Figura 21. Formato del archivo de topología para la definición de la prolina.** La primera línea del código (*RESI*) indica que a continuación se declarará un residuo, *PRO*, el cuál presenta una carga total de *0.00*. Posteriormente, bajo la instrucción *ATOM* se listan todos los átomos en el residuo con su nombre (*N*, *CD*, *HD1*), tipo (*N*, *CP3*, *HA*) y carga parcial (*-0.29*, *0.00*, *0.09*). La conectividad se define con el comando *BOND* y *DOUBLE*. La planaridad de los enlaces peptídicos está controlada bajo los comandos de torsiones no adecuadas (*IMPR*, de *impropers*). La instrucción *ACCEPTOR* sugiere los pares de átomos elegibles para formar puentes de hidrógeno. Finalmente, bajo el comando *IC* se calculan las distancias, ángulos y ángulos diedros (coordenadas internas) en segmentos de a cuatro átomos hasta abarcar la totalidad del aminoácido.

#### 4.1.4.1. Estados de protonación de la histidina

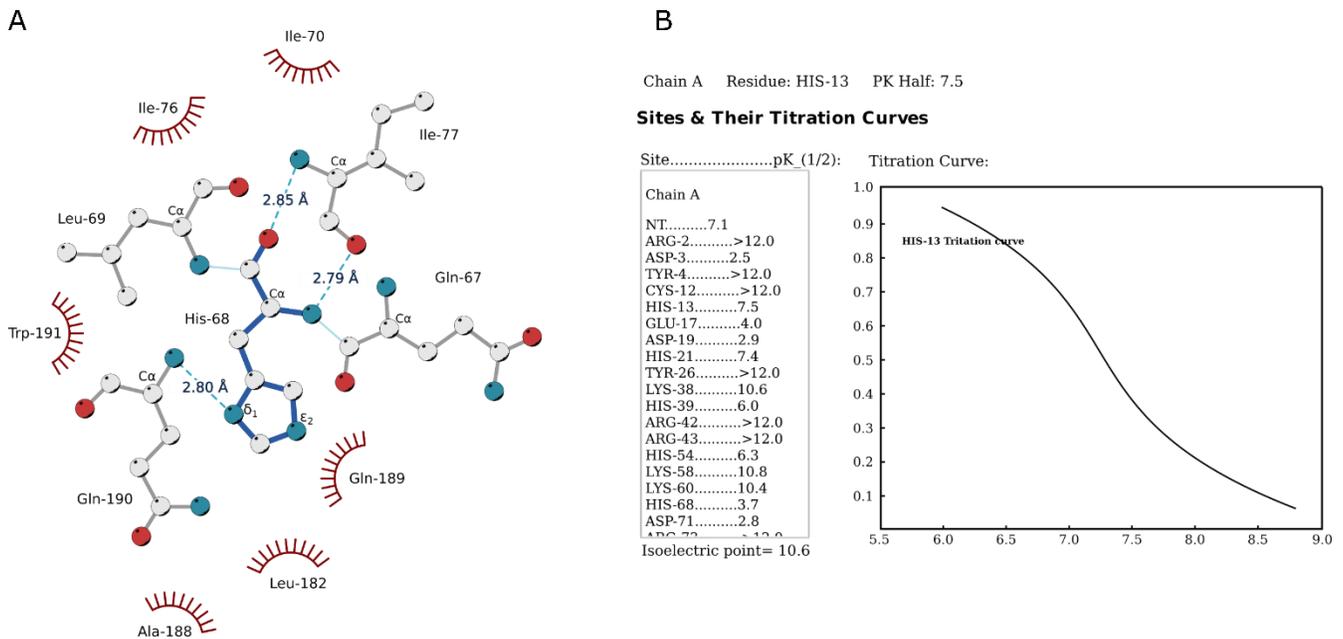
Los posibles estados de protonación de la histidina son un factor importante en las simulaciones por Dinámica Molecular por lo que deben estar propiamente definidos en el archivo de coordenadas con el fin de ser interpretadas por el archivo de topología. En la

Figura 22 se observa que la histidina puede estar protonada en el nitrógeno delta 1 ( $\delta_1$ , HSD), el nitrógeno épsilon 2 ( $\epsilon_2$ , HSE) o doblemente protonada y con carga neta positiva (HSP). En la mayoría de las estructuras cristalográficas es difícil determinar el estado de protonación de la histidina por lo que esto debe inferirse mediante estimaciones del pKa en combinación con análisis del entorno molecular de cada histidina.



**Figura 22. Posibles estados de protonación de la histidina.** Un valor de 6.04 para el pKa de la cadena lateral de la histidina permite tres estados de protonación en su anillo imidazol según si el protón está en el nitrógeno  $\delta_1$  (HSD),  $\epsilon_2$ , (HSE) o en ambos (HSP).

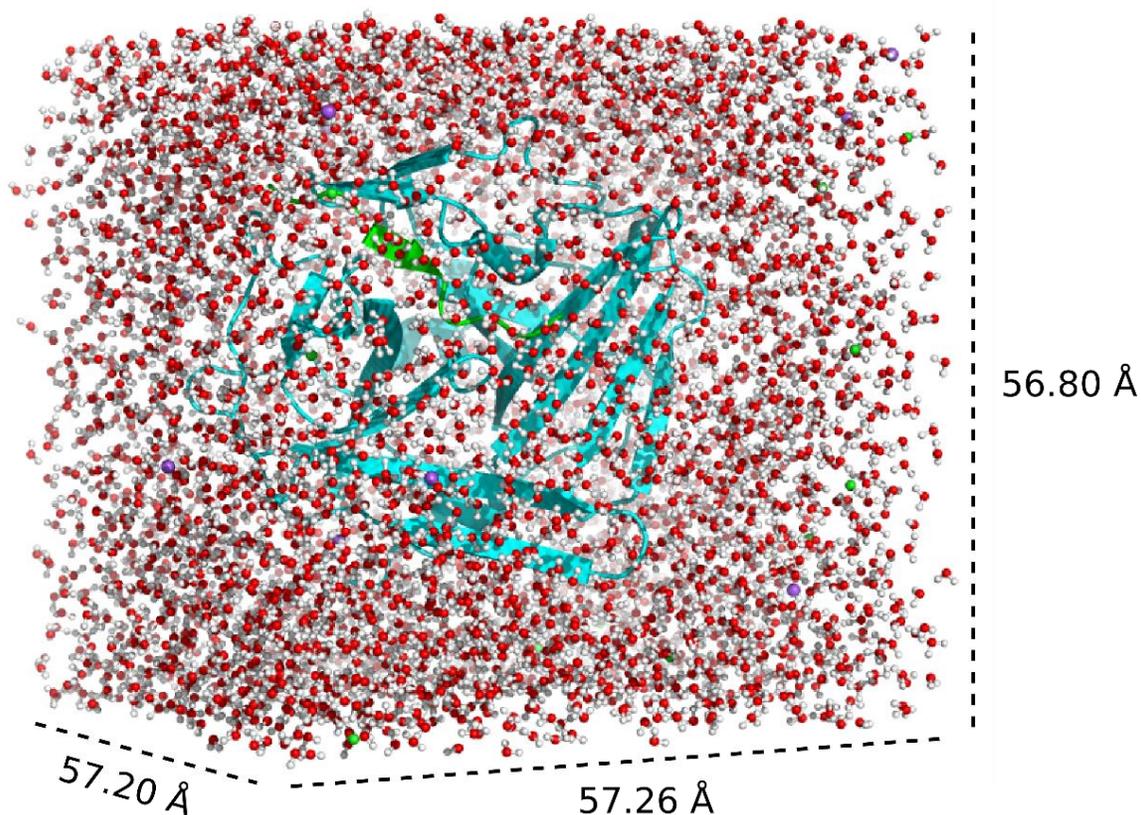
Para la predicción del pKa de cada histidina se utilizó H++ (*Virginia Tech*), un servidor que calcula los valores de pK de los grupos ionizables presentes en macromoléculas de acuerdo al pH del ambiente (Anandakrishnan *et al.*, 2012). Su aplicación permite predecir los estados de protonación de la histidina según el valor del pKa<sub>(1/2)</sub>, que por definición corresponde al pH en el cuál los estados de protonación tienen una posibilidad de ocurrir igual a 0.5. De tal forma que cuando los valores aportados por el software (Figura 23-B) sean inferiores al pH del medio, la histidina estará desprotonada y sus estados podrán ser HSD o HSE; cada uno con una probabilidad de ocurrir del 50% (Onufriev *et al.*, 2001). Para discriminar entre la forma HSD o HSE se analizó el entorno molecular utilizando el programa LigPlot, que permite visualizar la formación de puentes de hidrógeno cerca de la vecindad del anillo imidazol de la histidina, tal como se muestra en la Figura 23-A. Para el caso de TEV se encontró que tres histidinas (His-39, His-135, His-207) están protonadas en  $\delta_1$ , cuatro (His-13, His-54, His-68, His-160) están protonadas en  $\epsilon_2$ , y una (His-21) doblemente protonada. La histidina 39 de la tríada catalítica se modeló con protonación  $\delta_1$  (HSD).



**Figura 23. Definición de los estados de protonación de la histidina. A)** La protonación del anillo imidazol de la histidina 68 (resaltada en azul) ocurre en el nitrógeno épsilon 2 debido a la presencia de un puente de hidrógeno en el nitrógeno  $\delta_1$ . Por tanto, su nomenclatura en el archivo PDB cambia por HSE. **B)** Salida del servidor H++ (*Virginia Tech*). Los residuos de histidina (cadenas laterales protonables) que presentan valores de  $pK_{a(1/2)}$  inferiores al pH fisiológico ( $\sim 7.4$ ) sugieren un estado desprotonado HSE o HSD. La elección por uno u otro dependerá de factores como la orientación que tengan frente al solvente y puede ser determinada mediante software de visualización molecular. En este caso, nótese que la His-13 tiene un  $pK_{a(1/2)}$  mayor a 7.0, por lo que su nomenclatura en el archivo de coordenadas deberá cambiarse por HSP.

#### 4.1.4.2. Caja de solvatación

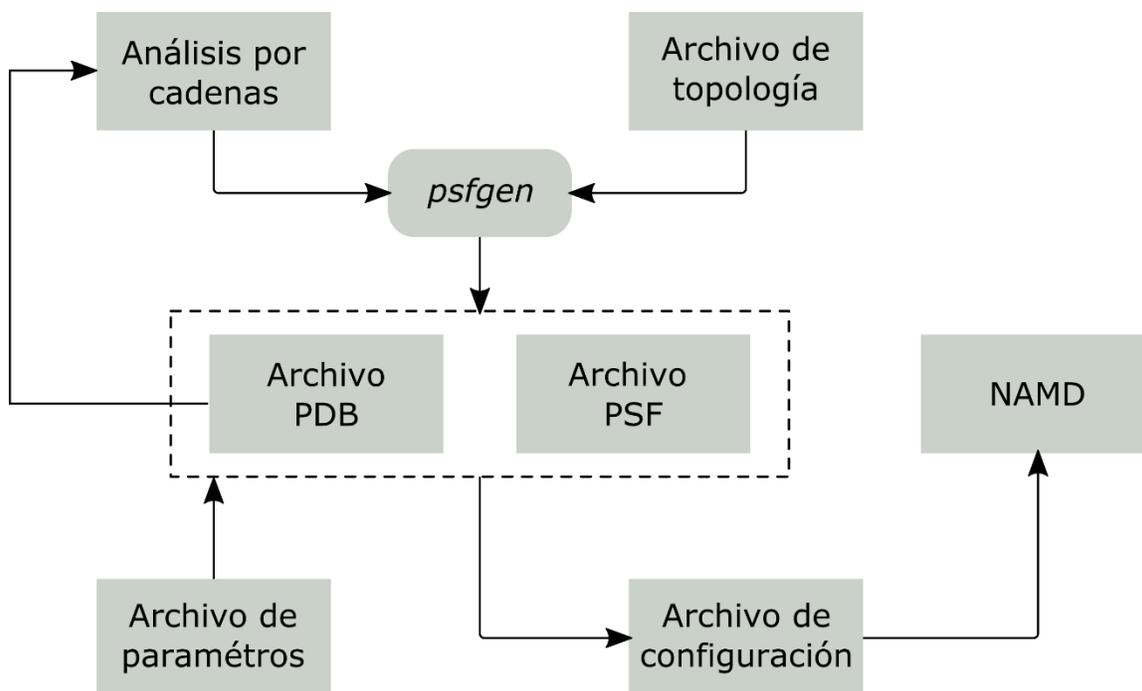
Una vez establecidos los estados de protonación de los grupos ionizables de la histidina, es posible generar el archivo de estructuras (PSF) y proceder a la solvatación de la proteína para representar más apropiadamente las condiciones de un ambiente celular. Para esto pueden emplearse dos metodologías que rodean la proteína: una esfera en el vacío o un cubo con condiciones periódicas de frontera. Esta última opción es la más aceptada y frecuente en simulaciones moleculares al eliminar el efecto frontera de las moléculas de agua cercanas a la superficie. En la Figura 24 se muestra la caja de solvatación creada con la extensión de VMD. La simulación incluye iones de sodio y cloro para garantizar la neutralidad del sistema.



**Figura 24. Caja de solvatación.** La proteasa Nla-Pro de TEV (cian) en complejo con el sustrato (verde) simulada en un ambiente acuoso. La concentración de iones de sodio y cloro es de 0.15 M y se representan mediante esferas de color violeta y verde, respectivamente.

#### 4.1.4.3. Archivo de configuración

La simulación en NAMD se ejecuta a través del archivo de configuración (*.conf*) donde se especifican los parámetros de temperatura, presión, reporte de datos de energía, condiciones periódicas de frontera y tiempos de simulación; también contiene información sobre la ubicación de los directorios con los archivos de entrada necesarios. En la Figura 25 se presenta un diagrama de flujo que resume cuáles son los requisitos para iniciar una simulación por Dinámica Molecular utilizando NAMD.



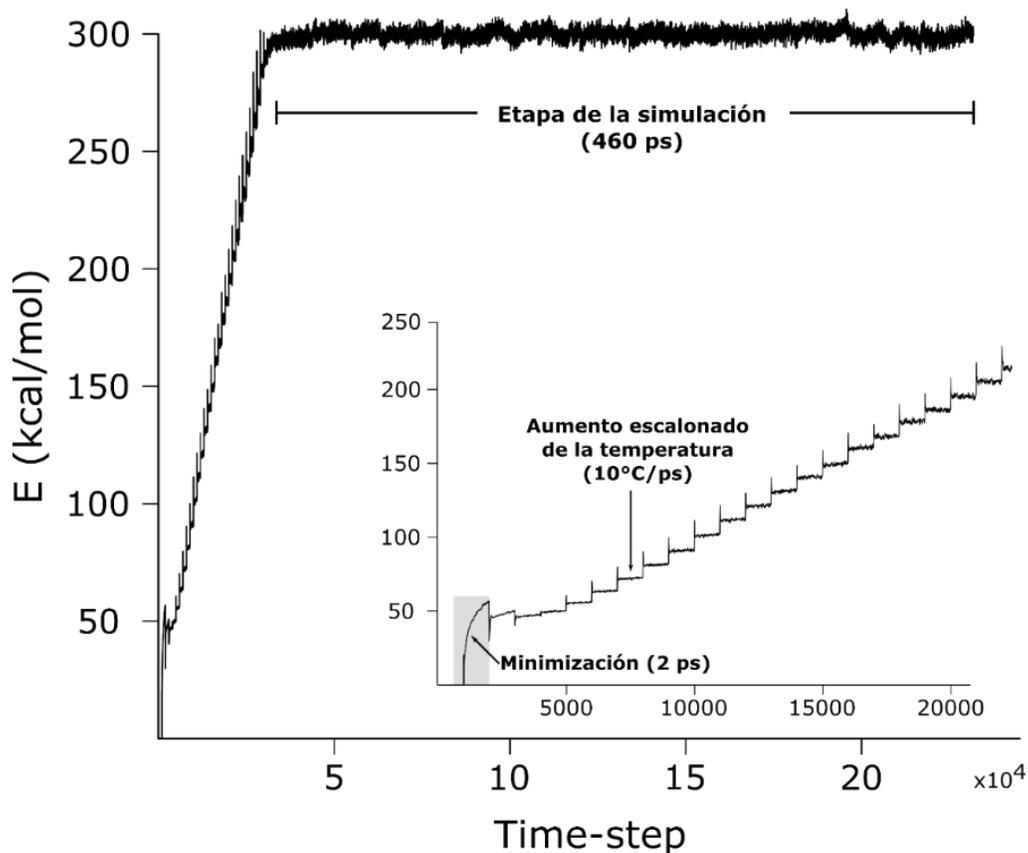
**Figura 25. Diagrama de flujo para las simulaciones por MD en NAMD.** Los archivos principales de entrada: coordenadas (PDB), estructura (PSF), topología (top\_all27\_prot\_lipid.inp) y parámetros (par\_all27\_prot\_lipid.inp), son necesarios en el archivo de configuración. La función *psfgen* corresponde a una extensión del software VMD.

La Figura 26 muestra las secciones más importantes del archivo de configuración utilizado para simular la Nla-Pro de TEV. Las condiciones periódicas de frontera se establecen mediante tres vectores ortogonales (*cellBasisVector 1, 2 y 3*) que definen una caja en tres dimensiones. El centro de la caja se define con el comando *cellOrigin*. Las condiciones de presión se definen con la instrucción *useGroupPressure*, calculando la presión del sistema según las fuerzas entre átomos y sus energías cinéticas. Al final del archivo se especifican los tiempos de equilibración y simulación del sistema. Las interacciones electrostáticas son calculadas mediante la metodología PME (*Particle Mesh Ewald*). Esta técnica acelera considerablemente el cálculo de las interacciones de Coulomb de alto rango que se presentan en sistemas de macromoléculas, mediante la sumatoria de dos series convergentes (suma de Ewald). Su fundamento físico consta en ubicar una carga puntual en una red (o cuadrícula) y rodearla de una nube iónica que apantalla las cargas vecinas, haciendo que las interacciones puedan ser consideradas como de corto alcance.

structure	../9_TEV_ionized.psf	←	Archivo PSF (estructuras)
coordinates	../9_TEV_ionized.pdb	←	Archivo PDB (coordenadas)
set outputname	9_TEV_final_298	←	Nombre archivo de salida
set temp_step	10	←	Aumento de la
firsttimestep	0		temperatura cada 10
paraTypeCharmm	on		Ubicación del archivo
parameters	../par_all27_prot_lipid.inp	←	de parámetros del
			campo de fuerza
temperature	\$temp_step		
if {1} {			Dimensiones de la caja de
cellBasisVector1	57.26 0 0	←	solvatación (condiciones
cellBasisVector2	0 57.21 0		periódicas de frontera)
cellBasisVector3	0 0 56.81		
cellOrigin	42.12 62.45 28.35	←	Centro de la caja
}			
wrapWater	on		
wrapAll	on		
timestep	2	←	Tiempo de paso definido en 2 fs
rigidBonds	all		
nonbondedFreq	1		Las moléculas en la
fullElectFrequency	2	←	simulación, especialmente las
stepspercycle	10		aguas, se definen como
			rígidas
if {1} {			La temperatura aumenta hasta
reassignTemp	\$temp_step		los 298 K. En cada incremento
reassignIncr	\$temp_step	←	la simulación continúa
reassignHold	298		corriendo durante mil pasos
reassignFreq	1000		
}			
outputName	\$outputname		
restartfreq	500		Reporte y guardado de
dcdfreq	100		información en intervalos
xstFreq	100	←	regulares de 500 pasos (~1
outputEnergies	10		ps). Impresión de las energías
outputPressure	100		cada 10 pasos (~0.02 ps)
binaryoutput	no		
fixedAtoms	off		
minimize	1000	←	1000 pasos en la etapa de
			minimización
run	230800	←	Número de pasos para la
			simulación

**Figura 26. Principales secciones del archivo de configuración.** Este archivo contiene las instrucciones ejecutadas en la simulación. Las explicaciones de algunos comandos se señalan por flechas.

El paso inicial de la simulación por Dinámica Molecular corresponde a la etapa de minimización (2 ps) en donde el campo de fuerza empleado inicia una búsqueda conformacional en todo el sistema hasta encontrar un mínimo de energía potencial (Figura 27). En este punto de estabilidad la fuerza neta sobre cada átomo es igual a cero, lo que permite dar paso al inicio de la simulación. Para esto, se inmoviliza la proteína en la caja de solvatación mientras que las moléculas de agua adoptan una orientación más estable. El proceso iterativo finaliza cuando el cambio de la energía respecto a las coordenadas de las moléculas sea cercano a cero.



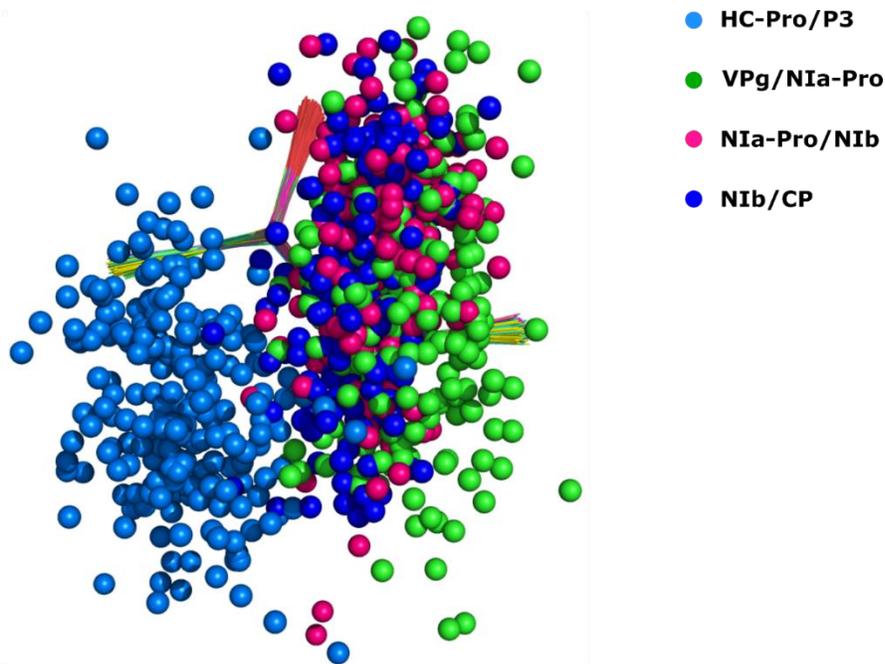
**Figura 27. Proceso de la simulación por Dinámica Molecular.** El paso inicial de minimización consiste en inmovilizar la proteasa para orientar las moléculas de agua y adecuar el sistema al paso del calentamiento escalonado (~6.71 ps), en donde el movimiento de la proteína se ha reactivado nuevamente. Cada escalón dura 0.22 ps en donde la temperatura aumenta 10 unidades hasta alcanzar 298 K. Una vez se logra la temperatura deseada, se procede a la etapa de simulación de 460 ps.

Posteriormente, se procede al paso de equilibración en donde a la proteína se le restaura su movimiento para que, de manera escalonada, el sistema alcance una temperatura de interés. En el proceso de calentamiento la temperatura aumenta 10 unidades cada 0.2 ps hasta

alcanzar 298 K, tomando un tiempo total de ~7 ps. Esta etapa requiere la asignación de velocidades en algún punto de baja temperatura (50 kcal/mol, Figura 27) para que luego, con la ecuación de movimiento en cada punto pueda determinarse la energía. Finalmente, la etapa de dinámica o producción es la de mayor interés pues es donde se determinan las propiedades termodinámicas del sistema y se extraen los datos de las diversas conformaciones adoptadas por la molécula en la simulación. En ella, se busca solucionar la segunda ecuación de movimiento de Newton en función de la energía potencial para cada átomo, utilizando métodos numéricos (diferencias finitas). Esta etapa abarcó un tiempo total de simulación de 460 ps.

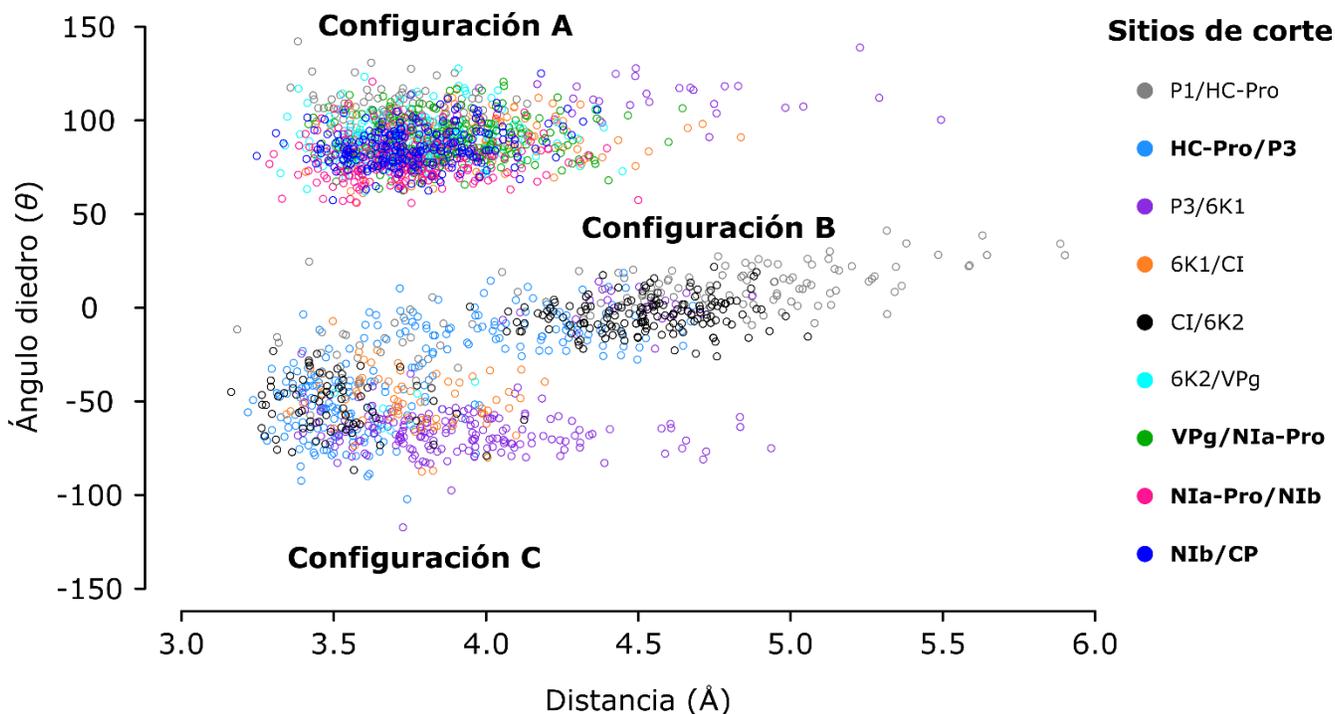
#### **4.1.5. Resultados de las simulaciones por Dinámica Molecular**

Con el fin de comparar cualitativamente las configuraciones que pueden adoptar los diferentes sitios de corte, se extrajeron las coordenadas de los átomos de azufre de la cisteína 144 (tríada catalítica) cada 0.02 ps de la simulación. De esta manera fue posible observar la posición relativa de la cisteína catalíticamente activa respecto al enlace peptídico a ser hidrolizado en diversos momentos de la simulación. La Figura 28 muestra la agrupación de tres sitios de corte (VPg/N1a-Pro, N1a-Pro/N1b y N1b/CP) incluyendo el sitio HC-Pro/P3 como control negativo. En esta figura es evidente que cuando N1a-Pro está unida a péptidos que contienen la secuencia de corte, los átomos de azufre responsables del ataque nucleofílico al carbono carbonílico se distribuyen en una región diferente respecto a la secuencia no específica (HC-Pro/P3). Lo anterior sugiere que los sitios de corte pueden diferenciarse en términos del procesamiento de acuerdo a la orientación, distancia y dirección de ataque del grupo tiol hacia el enlace peptídico blanco en el péptido-sustrato.



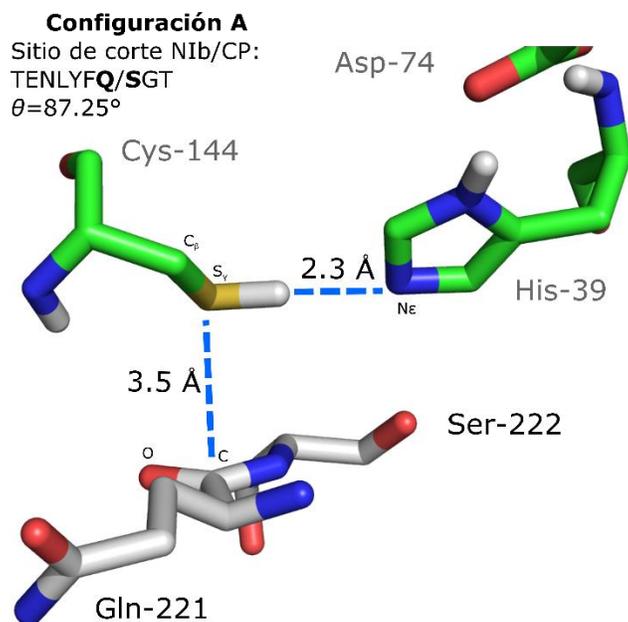
**Figura 28. Patrón de distribución de los átomos de azufre de la Cys-144 en diversos momentos de la simulación.** Los átomos de azufre de la cisteína 144 (tríada catalítica) se representan mediante esferas en códigos de color según el sitio de corte al que corresponden. Se observa que los sitios de corte específicos de la N1a-Pro comparten una misma región respecto al enlace peptídico a hidrolizarse. Por el contrario, HC-Pro/P3, cortado autocatalíticamente e independiente de N1a, se ubica en una zona más alejada. Lo anterior sugiere que la proteasa presenta diferencias en el procesamiento de acuerdo a la secuencia de los sustratos de corte.

Para facilitar el análisis de los datos, se evaluó la orientación del sustrato respecto a la cisteína de la tríada catalítica en cada sitio de corte midiendo la distancia entre el átomo de azufre (Cys-144) y el carbono carbonílico del enlace peptídico a ser hidrolizado además del ángulo diedro ( $\theta$ ) entre los planos definidos por el grupo tiol y el grupo carbonilo. Estos dos valores se tomaron cada 0.02 ps de la simulación dando un total de 232 datos. La Figura 29 presenta el patrón de distribución obtenido por región de corte. La distribución de valores para todas las simulaciones permite observar que la mayoría de los datos se concentran en tres configuraciones principales (A, B y C).



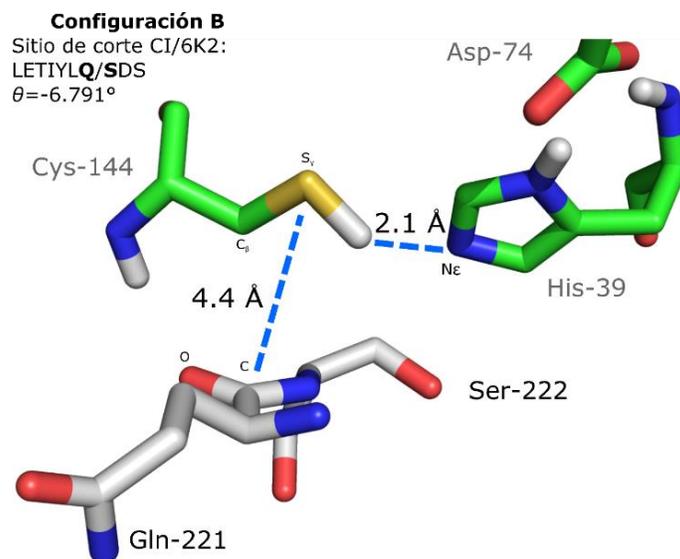
**Figura 29. Distribución de los sitios de corte simulados.** Cada uno de los sitios de corte se representa mediante datos bivariados de distancia (eje de las abscisas) y ángulo diedro (eje de las ordenadas). Se observa la presencia de tres regiones donde se agrupan la mayoría de los valores medidos y corresponden a las configuraciones A, B y C de mayor, media y poca probabilidad de corte, respectivamente.

En la Figura 30 se ilustra la configuración A donde se agrupan, en su mayoría, los sitios de mayor especificidad por la proteasa. Esta configuración presenta una geometría óptima de corte para la formación del complejo acil-enzima. La distancia entre el azufre del grupo tiol (Cys-144) y el nitrógeno épsilon,  $N\epsilon$  (His-39), es alrededor de 2.3 Å facilitando la protonación de la histidina mientras que la distancia entre el azufre y el carbono del enlace peptídico es de 3.5 Å, una distancia lo suficientemente corta como para facilitar el ataque nucleofílico. Finalmente, se observa un valor cercano a 90° para el ángulo diedro que corresponde a la geometría ideal para llevar a cabo la transición de una hibridación  $sp^2$  a  $sp^3$  en la formación del primer intermediario tetraédrico.



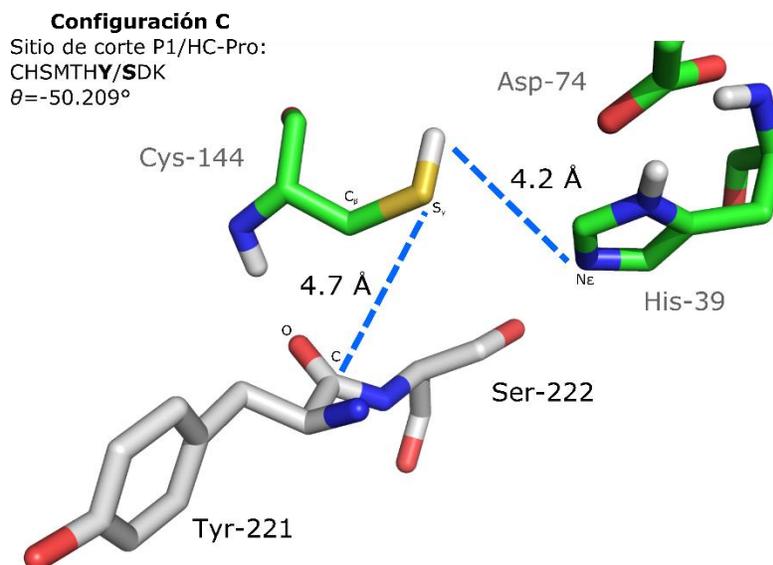
**Figura 30. Configuración A, tomando como referencia el corte Nib/CP.** Los residuos coloreados por átomos en verde corresponden a la tríada catalítica. En blanco se ilustran los residuos del sustrato entre los que ocurre el corte (carbono carbonílico nombrado como C). Se observa que el grupo tiol de la Cys-144 se encuentra perpendicular al carbono del enlace peptídico. Este factor y la corta distancia entre los átomos  $S_\gamma$  y C favorecen el ataque nucleofílico de la cisteína.

En contraste con la configuración A, la configuración B presenta un giro del enlace  $C_\beta-S_\gamma$  que aleja la posición del azufre a 4.4 Å del carbono del enlace peptídico (Figura 31). Si bien la distancia entre el hidrógeno y  $N_\epsilon$  (His-39) es cercana (2.1 Å), la ubicación más alejada del azufre no favorece el ataque nucleofílico y probablemente dificulta la formación del primer intermediario tetraédrico.



**Figura 31. Configuración B.** Los residuos coloreados por átomos en verde corresponden a la tríada catalítica. En blanco se ilustran los residuos del sustrato entre los que ocurre el corte. La distancia entre los átomos  $S_\gamma$  y  $C_\beta$  es mucho mayor que en la configuración A debido a una rotación en el enlace de los átomos  $C_\beta$  y  $S_\gamma$ . Consecuentemente, la orientación entre estos dos planos resulta desfasada, lo que termina en un proceso poco óptimo para el corte.

Finalmente, la configuración C (Figura 32) presenta un rotación del enlace  $C_\beta-S_\gamma$  que aleja el átomo de azufre del carbono del enlace peptídico una distancia de 4.7 Å, al igual que lo observado en la configuración B. Además, el grupo tiol presenta una rotación respecto al anillo imidazol de la His-39 que separa al hidrógeno del átomo  $N_\epsilon$  una distancia de 4.2 Å, lo que dificultaría la toma del protón por parte de la histidina y así volver el azufre un mejor nucleófilo. La configuración C es la más desfavorecida estructuralmente para llevar a cabo el proceso de hidrólisis del enlace peptídico.



**Figura 32. Configuración C.** Los residuos coloreados por átomos en verde corresponden a la tríada catalítica. En blanco se ilustran los residuos del sustrato entre los que ocurre el corte. Tanto la distancia entre los átomos  $S_\gamma$  y C, como entre el hidrógeno del grupo tiol y  $N_\epsilon$ , son mayores que en los casos predecesores sugiriendo una geometría molecular no propicia para el procesamiento del sustrato.

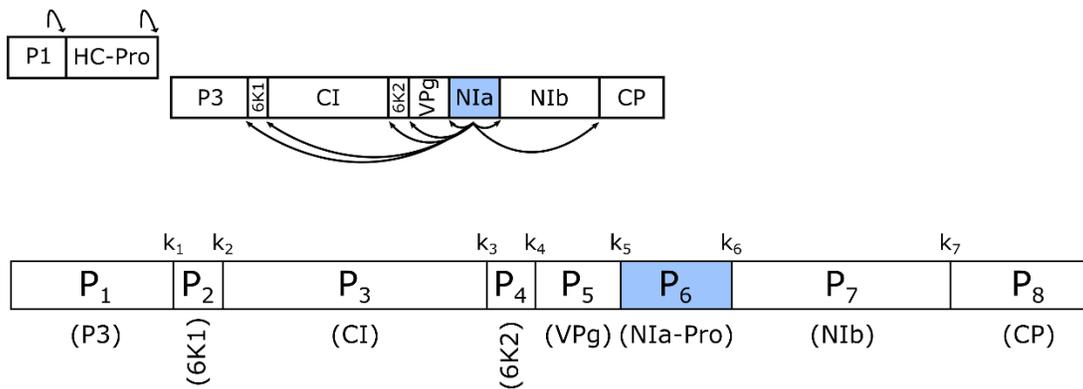
Estos resultados sugieren que el proceso de regulación proteolítica puede ser estudiado estructuralmente observando la orientación del grupo tiol de la cisteína respecto a la histidina de la triada catalítica y el sitio de corte en el sustrato. La presencia de diferentes aminoácidos en posiciones clave para la afinidad de la enzima y la geometría del entorno molecular en el sitio catalítico son factores que contribuyen al mecanismo de control del procesamiento de la poliproteína, ya que establecen un patrón de discriminación delimitado por escenarios de alta, media y poca efectividad de corte (Figura 29). Con el fin de usar la información de las simulaciones moleculares en la estimación de la tasa de procesamiento para las diferentes secuencias, se puede asumir que la velocidad de corte está directamente relacionada con la fracción del tiempo que pase el complejo enzima-sustrato en la configuración A. En este caso, la información obtenida por MD puede ser utilizada en un modelo cinético del proceso de maduración del virus como el que se discute a continuación.

#### 4.1.6. Modelo cinético

La proteasa Nla-Pro es responsable de la hidrólisis de los enlaces peptídicos en siete de los nueve sitios en la poliproteína. Para estudiar el proceso de maduración mediado por Nla-Pro

es necesario por lo tanto considerar siete constantes de velocidad y una secuencia de corte que inicia con un precursor que comprende la proteína P3 hasta CP (Figura 33). P1 y HC-Pro no hacen parte de este modelo ya que estas proteínas son removidas autocatalíticamente probablemente antes de la síntesis del resto de la poliproteína. La nomenclatura de los intermediarios empleada se ilustra en la Figura 33 y se define según la expresión:  $P_{ij}$ , donde  $i$  toma valores del 1 al 7 representando el péptido de inicio desde el extremo C-terminal; y  $j$  toma valores del 2 al 8, constituyendo el péptido final hacia el extremo N-terminal.

## Poliproteína



**Figura 33. Nomenclatura usada en el modelo enzimático.** En el panel inferior se muestra la poliproteína de ocho constitutivos que es procesada por Nla-Pro. Las constantes se ubican arriba de sus respectivos sitios de corte ( $k_1, k_2, \dots, k_7$ ). La nomenclatura de los precursores depende del rango de péptidos que contenga. Por ejemplo, el precursor de las proteínas P3, 6K1, CI y 6K2 se denomina  $P_{14}$ ; el precursor de VPg, Nla-Pro y NIb se denomina  $P_{57}$ .

El procesamiento proteolítico de la poliproteína a partir del precursor  $P_{18}$  (P3-CP) genera un máximo de 36 especies moleculares que luego de un tiempo suficientemente grande debe resultar en la formación de las proteínas maduras P3, 6K1, CI, 6K2, VPg, Nla-Pro, NIb y CP en cantidades equimolares ( $P_1, P_2, P_3, P_4, P_5, P_6, P_7$  y  $P_8$ , respectivamente, según la nomenclatura descrita previamente). El modelo propuesto en este trabajo asume que el corte en cualquiera de los siete sitios puede darse de manera intra o intermolecular dependiendo de que el precursor contenga o no el dominio Nla-Pro ( $P_6$ ).

La relación entre constantes intra e intermoleculares puede estimarse con la teoría del estado de transición donde las constantes de velocidad ( $k$ ) están relacionadas con la energía libre de activación mediante la ecuación 1:

$$k = \alpha \frac{k_B T}{h} e^{\frac{\Delta G}{RT}} = \alpha \frac{k_B T}{h} e^{\frac{\Delta H}{RT}} e^{\frac{-\Delta S}{R}}$$

**Ecuación 1**

Donde  $\alpha$  se conoce como la constante de transmisión y el factor  $k_B T/h$  corresponde a la frecuencia vibracional del enlace en el estado de transición definida por la temperatura (T), la constante de Boltzmann,  $k_B$ , y la constante de Planck,  $h$ . Según esto, es posible relacionar las constantes de velocidad entre ambos procesos de acuerdo a la ecuación 2:

$$\frac{k_{inter}}{k_{intra}} = \frac{e^{\frac{\Delta H_{inter}}{RT}} e^{\frac{-\Delta S_{inter}}{R}}}{e^{\frac{\Delta H_{intra}}{RT}} e^{\frac{-\Delta S_{intra}}{R}}}$$

**Ecuación 2**

El factor entálpico,  $\Delta H$ , depende de las interacciones moleculares responsables de la formación del complejo enzima-sustrato. Por lo tanto, este término puede ser cancelado al considerar que, tanto en las reacciones inter como intramoleculares, la proteasa involucra los mismos residuos en la secuencia de corte y suceden a una misma temperatura (298 K). La expresión queda reducida a la ecuación 3:

$$\frac{k_{inter}}{k_{intra}} = \frac{e^{\frac{-\Delta S_{inter}}{R}}}{e^{\frac{-\Delta S_{intra}}{R}}}$$

**Ecuación 3**

De la ecuación 3 se deduce que la principal diferencia entre las constantes se debe a un factor entrópico. En las reacciones químicas los cambios de entropía pueden separarse en términos de translación, rotación, vibración y rotación interna, como se expresa en la ecuación 4:

$$\frac{k_{\text{inter}}}{k_{\text{intra}}} = e^{\left(\frac{\Delta\Delta S}{R}\right)_{\text{trans}}} e^{\left(\frac{\Delta\Delta S}{R}\right)_{\text{rot}}} e^{\left(\frac{\Delta\Delta S}{R}\right)_{\text{vib}}} e^{\left(\frac{\Delta\Delta S}{R}\right)_{\text{rot-interna}}}$$

#### Ecuación 4

Donde  $\Delta\Delta S$  corresponde a la resta entre un término de entropía intermolecular y un término de entropía intramolecular. Al suponer que la entropía interna rotacional y vibracional pueden ser descartadas por su mínimo aporte a un complejo no covalente, el costo para inmovilizar una molécula entre las dos entropías restantes de rotación y translación sería del orden de  $44 \text{ cal}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$  (Page y Jencks, 1971).

$$k_{\text{inter}} = k_{\text{intra}} e^{\left(\frac{\Delta\Delta S}{R}\right)_{\text{trans}}} e^{\left(\frac{\Delta\Delta S}{R}\right)_{\text{rot}}} = k_{\text{intra}} \times A$$

#### Ecuación 5

El valor de A en la ecuación 5 es del orden de  $10^{-9}$  ( $4.13 \times 10^{-9} \text{ M}$ ) y debe ser considerado como un factor de corrección en el planteamiento de las tasas de reacción entre procesos intra e intermoleculares que ocurren en el procesamiento de la poliproteína.

Las ecuaciones de la cinética enzimática para el procesamiento de la poliproteína involucran 36 especies, incluyendo los péptidos finales. Dicho planteamiento se organizó en un sistema matricial de la forma  $E(t)\cdot x = x'$ .

$$\begin{array}{ccc}
 E(t) & x & x' \\
 \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} & \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ \vdots \\ k_n \end{bmatrix} & = \begin{bmatrix} \frac{d[E]_1}{dt} \\ \frac{d[E]_2}{dt} \\ \vdots \\ \frac{d[E]_m}{dt} \end{bmatrix}
 \end{array}$$

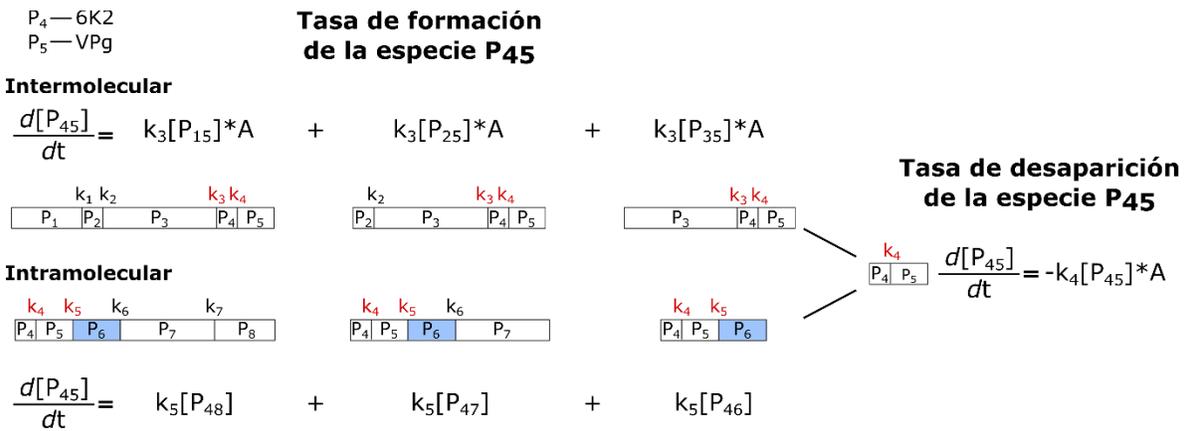
Donde  $E(t)$  es la matriz de coeficientes de la concentración de especies y precursores involucrados en el procesamiento,  $x$  es el vector de las variables que representa las siete constantes cinéticas y  $x'$  es el vector solución de dimensión  $36 \times 1$ ; el máximo valor para  $m$  y  $n$  es de 36 y 7, respectivamente. Un componente,  $a_{mn}$ , de la matriz  $E(t)$  (Figura 34) puede

$-P_{18}$	$-P_{18}$	$-P_{18}$	$-P_{18}$	$-P_{18}$	$-P_{18}$	$-P_{18}$
$-P_{17}$	$-P_{17}$	$-P_{17}$	$-P_{17}$	$-P_{17}$	$-P_{17}$	$P_{18}$
$P_{18}$	$-P_{28}$	$-P_{28}$	$-P_{28}$	$-P_{28}$	$-P_{28}$	$-P_{28}$
$-P_{16}$	$-P_{16}$	$-P_{16}$	$-P_{16}$	$-P_{16}$	$P_{18}+P_{17}$	0
$P_{17}$	$-P_{27}$	$-P_{27}$	$-P_{27}$	$-P_{27}$	$-P_{27}$	$P_{28}$
0	$P_{18}+P_{28}$	$-P_{38}$	$-P_{38}$	$-P_{38}$	$-P_{38}$	$-P_{38}$
$-P_{15}*A$	$-P_{15}*A$	$-P_{15}*A$	$-P_{15}*A$	$P_{18}+P_{17}+P_{16}$	0	0
$P_{16}$	$-P_{26}$	$-P_{26}$	$-P_{26}$	$-P_{26}$	$P_{27}+P_{28}$	0
0	$P_{17}+P_{27}$	$-P_{37}$	$-P_{37}$	$-P_{37}$	$-P_{37}$	$P_{38}$
0	0	$P_{18}+P_{28}+P_{38}$	$-P_{48}$	$-P_{48}$	$-P_{48}$	$-P_{48}$
$-P_{14}*A$	$-P_{14}*A$	$-P_{14}*A$	$P_{15}*A+P_{16}+P_{17}+P_{18}$	0	0	0
$P_{15}*A$	$-P_{25}*A$	$-P_{25}*A$	$-P_{25}*A$	$P_{28}+P_{27}+P_{26}$	0	0
0	$P_{16}+P_{26}$	$-P_{36}$	$-P_{36}$	$-P_{36}$	$P_{38}+P_{37}$	0
0	0	$P_{17}+P_{27}+P_{37}$	$-P_{47}$	$-P_{47}$	$-P_{47}$	$P_{48}$
$-P_{13}*A$	$-P_{38}*A$	$P_{18}+P_{17}+P_{16}+P_{15}*A+P_{14}*A$	0	0	0	0
$P_{14}*A$	$-P_{24}*A$	$-P_{24}*A$	$P_{28}+P_{27}+P_{26}+P_{25}*A$	0	0	0
0	$P_{15}*A+P_{25}*A$	$-P_{35}*A$	$-P_{35}*A$	$P_{38}+P_{37}+P_{36}$	0	0
0	0	$P_{16}+P_{26}+P_{36}$	$-P_{46}$	$-P_{46}$	$P_{48}+P_{47}$	0
0	0	0	$P_{17}+P_{27}+P_{37}+P_{47}$	$-P_{57}$	$-P_{57}$	$P_{58}$
$-P_{12}*A$	$P_{18}+P_{17}+P_{16}+P_{15}*A+P_{14}*A+P_{13}*A$	0	0	$P_{18}+P_{28}+P_{38}+P_{48}+P_{58}$	$-P_{68}$	$-P_{68}$
$P_{13}*A$	$-P_{23}*A$	$P_{28}+P_{27}+P_{26}+P_{25}*A+P_{24}*A$	0	0	0	0
0	$P_{14}*A+P_{24}*A$	$-P_{34}*A$	$P_{38}+P_{37}+P_{36}+P_{35}*A$	0	0	0
0	0	$P_{15}*A+P_{25}*A+P_{35}*A$	$-P_{45}*A$	$P_{48}+P_{47}+P_{46}$	0	0
0	0	$P_{15}*A+P_{25}*A+P_{35}*A$	$-P_{45}*A$	$P_{48}+P_{47}+P_{46}$	0	0
0	0	0	$P_{16}+P_{26}+P_{36}+P_{46}$	$-P_{56}$	$P_{58}+P_{57}$	0
0	0	0	0	$P_{17}+P_{27}+P_{37}+P_{47}+P_{57}$	$-P_{67}$	$P_{68}$
0	0	0	0	0	$P_{18}+P_{28}+P_{38}+P_{48}+P_{58}+P_{68}$	$-P_{78}*A$
$P_{12}*A$	$P_{28}+P_{27}+P_{26}+P_{25}*A+P_{24}*A+P_{23}*A$	0	0	0	0	0
$P_{18}+P_{17}+P_{16}+P_{15}*A+P_{14}*A+P_{13}*A+P_{12}*A$	0	$P_{38}+P_{37}+P_{36}+P_{35}*A+P_{34}*A$	0	0	0	0
0	$P_{13}*A+P_{23}*A$	$P_{14}*A+P_{24}*A+P_{34}*A$	$P_{48}+P_{47}+P_{46}+P_{45}*A$	0	0	0
0	0	0	$P_{15}*A+P_{25}*A+P_{35}*A+P_{45}*A$	$P_{58}+P_{57}+P_{56}$	0	0
0	0	0	0	$P_{16}+P_{26}+P_{36}+P_{46}+P_{56}$	$P_{68}+P_{67}$	0
0	0	0	0	0	$P_{17}+P_{27}+P_{37}+P_{47}+P_{57}+P_{67}$	$P_{78}*A$
0	0	0	0	0	0	$P_{18}+P_{28}+P_{38}+P_{48}+P_{58}+P_{68}+P_{78}*A$

**Figura 34. Matriz de especies  $E(t)$ .** Algunos componentes de la matriz representan la sumatoria de todos los precursores involucrados en la tasa de formación y desaparición de la concentración de la especie correspondiente. El factor A es la corrección por entropía que requiere inmovilizar una molécula en un proceso intermolecular. La cinética para la especie sombreada se explica en detalle en la Figura 35.

abarcar diversos precursores que intervienen en las tasas de producción y desaparición de la especie. Por lo tanto, el término puede ser expresado como la sumatoria de las concentraciones de las especies partícipes.

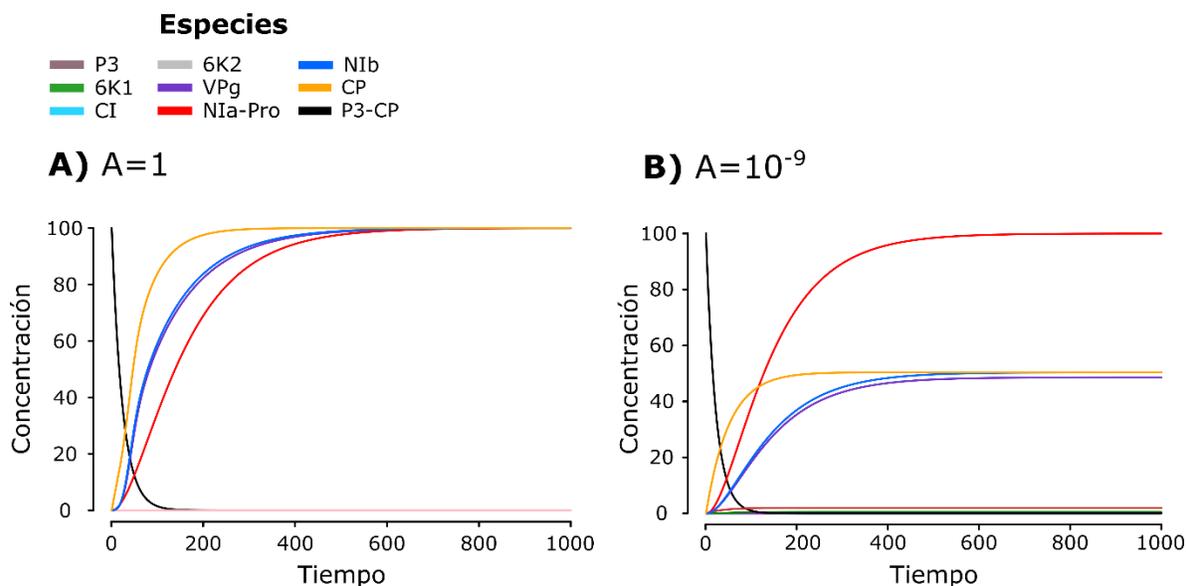
Según la Figura 35, la tasa de formación del dipéptido P<sub>45</sub> involucra un evento inter e intramolecular en los que deben ser procesados los precursores P<sub>15</sub>, P<sub>25</sub>, P<sub>35</sub>, P<sub>48</sub>, P<sub>47</sub>, P<sub>46</sub>. Una vez formado, su tasa de desaparición permite obtener las proteínas maduras 6K2 y VPg.



**Figura 35. Cinética del dipéptido P<sub>45</sub> (6K2-VPg).** La tasa de formación de P<sub>45</sub> involucra un proceso inter e intramolecular. En el primer caso se observa la presencia del factor de corrección, A, en cada proceso donde Nla-Pro interviene para la generación del corte. Finalmente, P<sub>45</sub> también es procesado de manera intermolecular para obtener 6K2 y VPg. La proteasa (P<sub>6</sub>) Nla-Pro se señala en azul y las constantes cinéticas que participan en la cinética del proceso se resaltan en rojo.

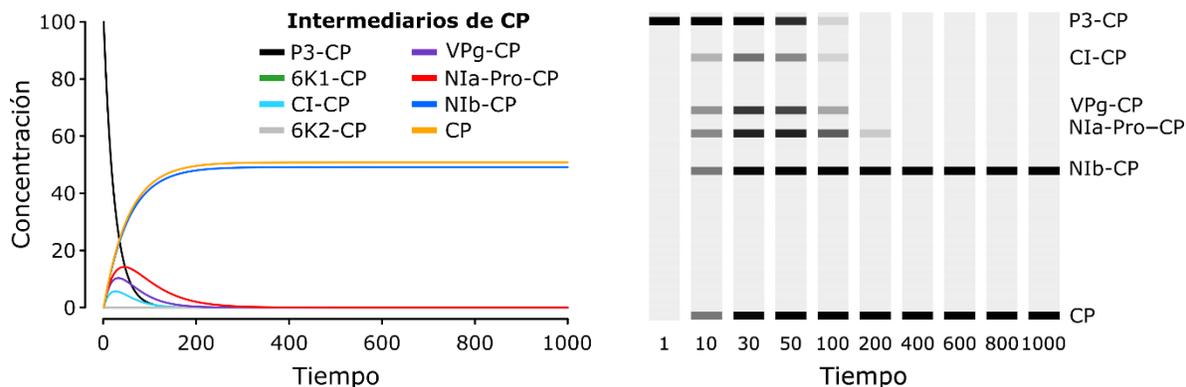
El efecto de la entropía en el proceso de maduración se ilustra en la Figura 36. Se evidencia que las tasas de producción fueron diferentes entre dos escenarios: cuando A=1, las concentraciones de las especies maduras son mucho mayores y sus tasas de formación son más rápidas. El primer péptido en aparecer es CP, seguido de la replicasa, el domino VPg y la proteasa (como producto maduro). Lo anterior sugiere que el procesamiento intermolecular por parte de la Nla-Pro no presenta inconvenientes de eficiencia para la síntesis de proteínas maduras. Por otro lado, cuando A=10<sup>-9</sup> se observa que las concentraciones finales de las proteínas se reducen considerablemente respecto al caso anterior. Solo Nla-Pro presenta una concentración similar y es la segunda en ser madurada, superada

por CP; la replicasa se obtiene en tercer lugar, seguido de VPg. Lo anterior permite concluir que, en el modelo enzimático propuesto, sí hay evidencia del mecanismo intermolecuar llevado a cabo por NIa-Pro en el procesamiento de la poliproteína, cuya influencia recae principalmente en las concentraciones de los productos finales.



**Figura 36. Cinética de maduración en escenarios inter e intramoleculares. A)** Al no tener en cuenta un factor de corrección para los procesos intermoleculares, la concentración de las especies maduras es 2 veces mayor que en **B)** un escenario donde la entropía sí es considerada en eventos de corte intermoleculares.

Los resultados de la cinética también pueden ser presentados mediante la simulación de pruebas de Western blot para la identificación de intermediarios que contengan la cápside (con un anticuerpo anti-CP). Como se observa en la Figura 37, el precursor P3-CP es procesado en 100 unidades de tiempo relativas, mientras se obtienen cuatro especies intermediarias con diferentes tiempos de duración: CI-CP y VPg-CP presentan los menores tiempos, seguido de NIa-Pro-CP; NIb-CP y la proteína madura CP alcanzan su concentración final aproximadamente en el mismo tiempo. Esta representación es útil a la hora de comparar los resultados de las simulaciones moleculares con experimentos realizados en TEV, PVA, TVMV, entre otros.



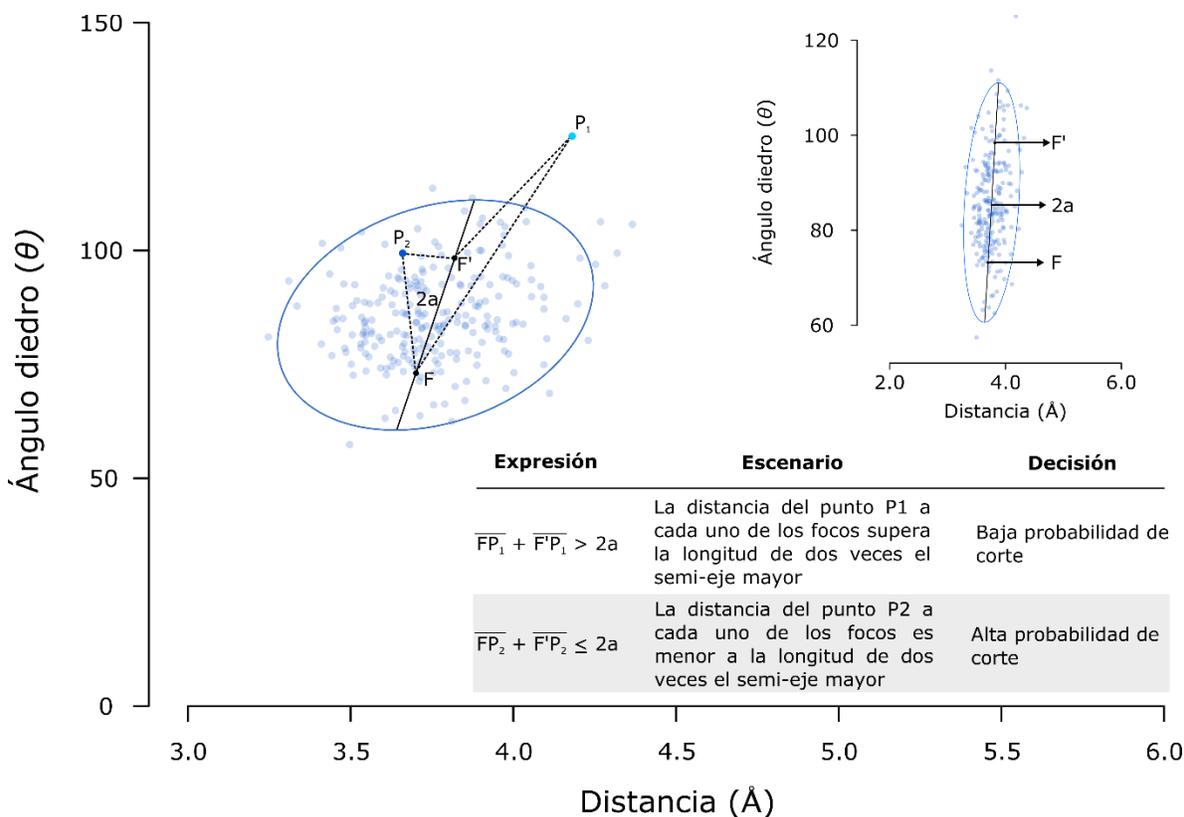
**Figura 37. Simulación de Western blot para especies con CP.** Los intermediarios que presentan CP en la cinética del procesamiento de la poliproteína son evidenciados en un resultado simulado de Western blot para anti-CP. Se destacan los diferentes tiempos entre especies y la rápida concentración final de la proteína de cápside madura.

#### 4.1.7. Análisis de las frecuencias de corte en la poliproteína de TEV

Partiendo de la suposición que N1b/CP conforma el sitio de corte más eficiente de la poliproteína, es posible estimar las tasas relativas de corte para cada uno de los otros sitios. Esto puede ser empleado en la evaluación de constantes cinéticas si se asume que la mayor probabilidad de corte depende de la frecuencia en que las moléculas presenten una orientación similar a la configuración A (Figura 30), es decir, una distancia del orden de 3.6 Å entre el azufre de la Cys-144 ( $S_V$ ) y el carbono carbonílico del enlace peptídico a ser hidrolizado (C); junto con un ángulo diedro cercano a 90° entre los planos del enlace tiol y el grupo carbonilo.

Para esto se estableció una frontera de decisión que delimita la región correspondiente a la configuración A utilizando el software estadístico R y datos de la simulación de TEV en complejo con el sitio de corte N1b/CP (Figura 38). La función *ellipse* permite definir la trayectoria de una elipse implementando la metodología descrita por Murdoch y Chow (1996). Los parámetros de entrada necesarios son: vectores de datos (*cor*) que corresponden a la distancia (abscisa) y ángulo diedro (ordenada) del corte de referencia N1b/CP, matriz de covarianza según la desviación estándar (*sd*) de estos valores de entrada, el centro de la

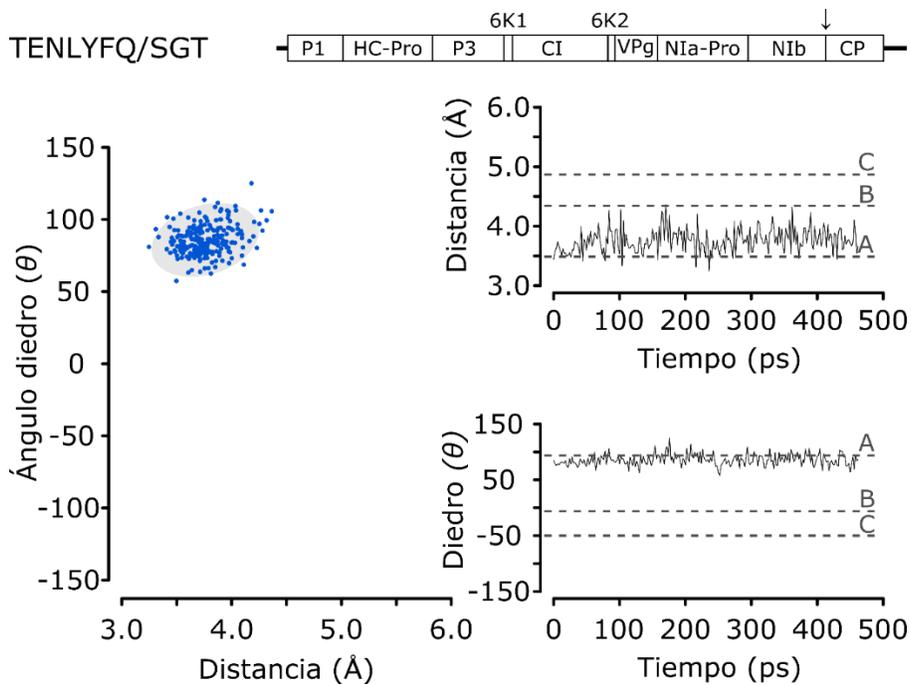
elipse según el promedio (*mean*) de cada uno de los datos y un límite de confianza (*level*). La tasa de corte está definida por la fracción de puntos que caen dentro de la región de decisión durante el tiempo de la simulación.



**Figura 38. Frontera de decisión según el corte Nlb/CP de TEV.** Los puntos azules de la gráfica corresponden a diversos momentos de la simulación extraídos cada 0.02 ps (232 en total). La probabilidad de efectuar un corte puede establecerse de acuerdo al número de puntos que caen en la sección definida por una elipse. Según sus propiedades, la suma de la distancia de cualquier punto definido en su dominio hacia sus dos focos ( $F$  y  $F'$ ), no debe superar una longitud equivalente a dos veces su semi-eje mayor ( $a$ ). En la esquina superior derecha se ilustra la misma elipse sin la distorsión causada por la escala de los ejes.

Según lo anterior, el complejo Nlb/CP estuvo en un 93.97% dentro de la zona definida como favorable al corte. Como se muestra en la Figura 39, la distancia entre átomos  $S_V$ -C estuvo en un promedio de 3.7 Å, mientras que para el ángulo diedro fue de 85.83°, sin la presencia de saltos o bruscas fluctuaciones a lo largo del tiempo de simulación (460 ps). Dicha región de corte corresponde al mejor escenario para lograr la separación de los péptidos (configuración A), pues presentó la distancia óptima entre el azufre de la cisteína y su correcta orientación

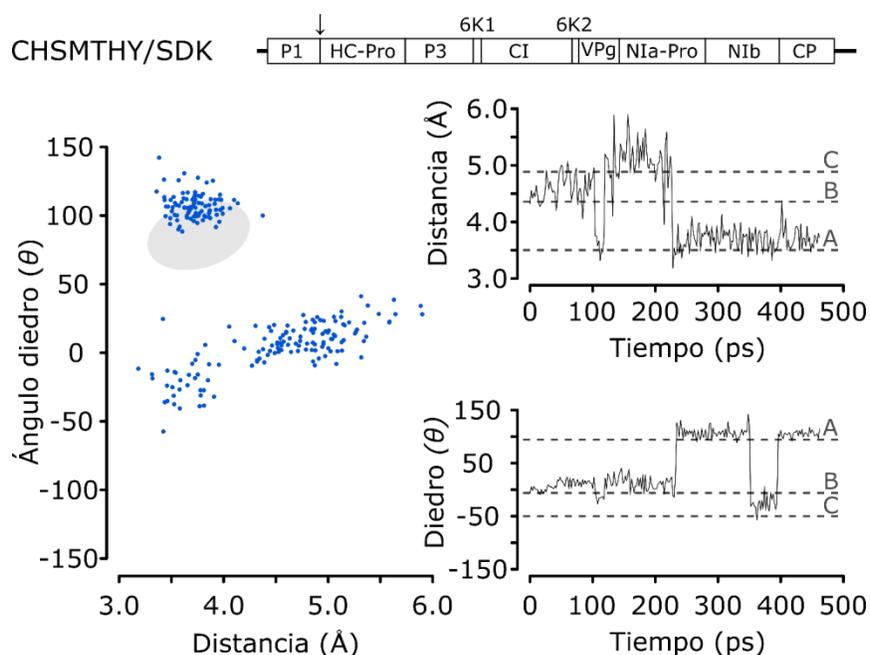
respecto al plano peptídico para dirigir el ataque nucleofílico hacia el carbono carbonílico, durante un máximo intervalo de tiempo. No se observó ningún salto hacia otras las regiones de las configuraciones B o C.



**Figura 39. Sitio de corte N1b/CP.** La concentración de puntos dentro de la elipse de decisión sugiere que el sustrato preservó un estado óptimo para su hidrólisis. Esta orientación se observó a lo largo de los 460 ps de la simulación y permite definir una elevada tasa de procesamiento de NIa-Pro sobre el sitio N1b/CP. Sobre el panel derecho se indican las posibles conformaciones de la molécula (líneas grises) y su preferencia por la configuración A. La frecuencia en la región de corte fue del 93.97%.

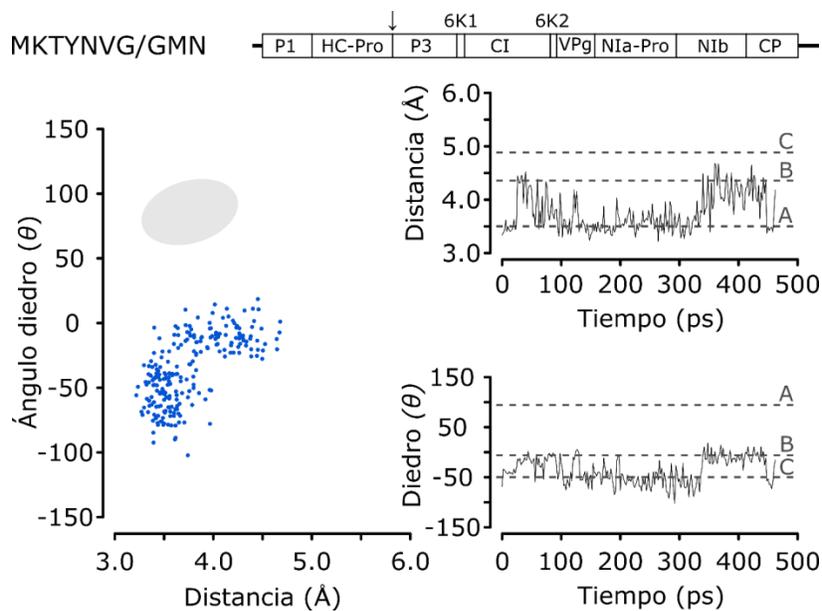
En contraposición a estos resultados, se encuentra el procesamiento de los sitios P1/HC-Pro y HC-Pro/P3 que no son procesados por la enzima NIa-Pro (Figura 40). En el primer caso, durante 200 ps, la configuración del péptido-sustrato no fue óptima para efectuar el corte pues se presentó una constante fluctuación en las distancias durante un valor del ángulo diedro cercano a 0°. A partir de dicho momento, la simulación mostró una distancia azufre-carbono óptima alrededor de 3.5 Å y un ángulo diedro cercano a 100°, presentando un tiempo de aproximadamente 300 ps en el que la orientación molecular de la proteasa NIa-Pro estuvo cerca de la configuración A del sitio activo. En este intervalo la

distancia entre átomos S<sub>V</sub>-C fue de 4.2 Å y el ángulo diedro de 45.006°, aportando una frecuencia de puntos en la región de corte de 22.84%.



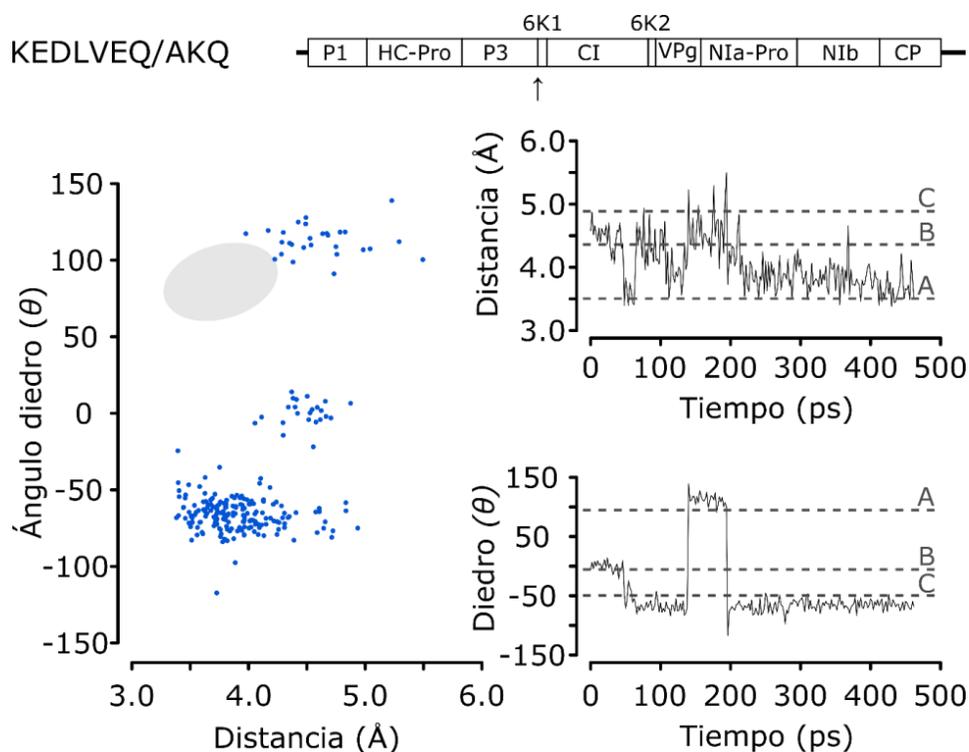
**Figura 40. Sitio de corte P1/HC-Pro.** Esta región, inespecífica de NIa-Pro, presentó una frecuencia de corte del 22.84%. La enzima y el sustrato adoptaron, por lo menos parcialmente, una orientación cercana a la configuración A. En la otra mitad del tiempo de simulación se observaron fluctuaciones entre las configuraciones B y C (líneas grises).

A diferencia del caso anterior, el péptido-sustrato de la región HC-Pro/P3 no presentó en ningún momento una configuración propensa al corte (Figura 41). La distancia entre átomos osciló entre 3.2 y 4.6 Å con un promedio de 3.8 Å. Por su parte, el ángulo diedro tuvo un valor promedio de 36°. Estos resultados son consistentes con la biología de los potyvirus ya que los sitios P1/HC-Pro y HC-Pro/P3 son procesados autocatalíticamente por las proteasas P1 y HC-Pro, independientes de NIa-Pro.



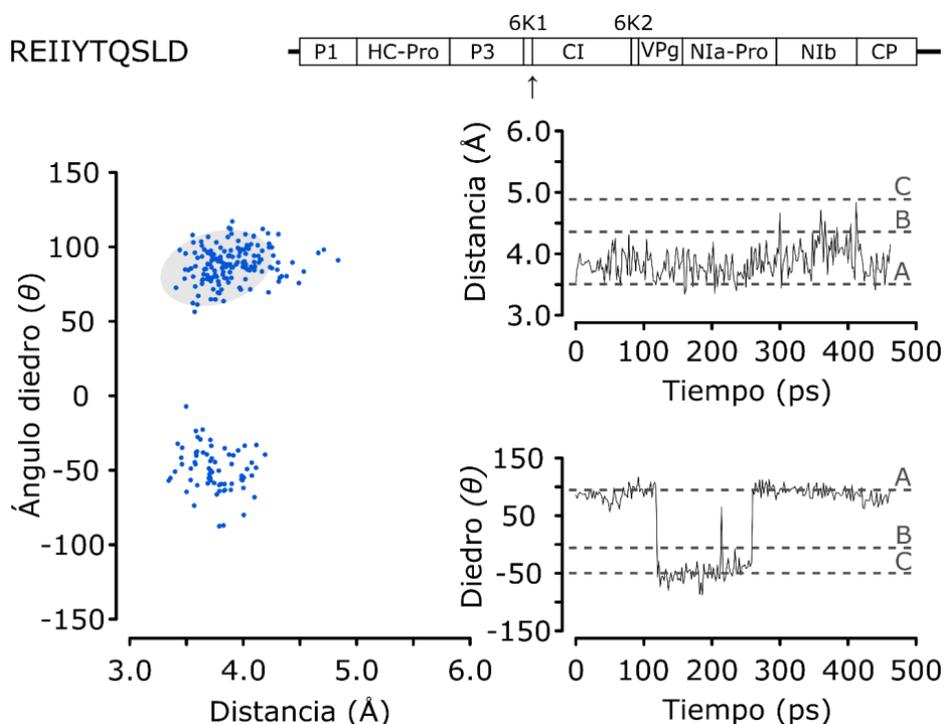
**Figura 41. Sitio de corte HC-Pro/P3.** Esta región es procesada autocatalíticamente por HC-Pro lo que resulta coherente con una frecuencia en la zona de corte del 0%. La distancia S<sub>γ</sub>-C presentó valores cercanos a la configuración A por diversos intervalos de tiempo de la simulación; sin embargo, la orientación del ángulo diedro estuvo entre las configuraciones B y C (líneas grises), haciendo menos probable la ocurrencia del corte.

El precursor P3/6K1 no alcanzó una configuración óptima de corte por NIa-Pro. La distancia entre el átomo de azufre y el carbono carbonílico tuvo un valor promedio de 4.045 Å durante los 460 ps de la simulación. Entre los 140 y 200 ps se logró adoptar un estado cercano a la configuración A que favorece el ataque nucleofílico entre el plano peptídico y el azufre de la cisteína (ángulo diedro promedio de 104.81°); sin embargo, la configuración no trascendió en la molécula y retornó rápidamente al estado precedente hasta finalizar la simulación aportando una frecuencia de puntos en la región de corte del 0% (Figura 42). Este sitio presenta una diferencia significativa respecto al sitio de corte de referencia NIb/CP (valina en lugar de tirosina en la posición P-3). El resultado es compatible con el reporte de Yoon y colaboradores, quienes no detectaron procesamiento del precursor P3/6K1 *in vitro* utilizando la proteasa NIa-Pro recombinante en *Tobacco vein mottling virus* (TVMV) (Yoon *et al.*, 1992). Esto indica que el corte del sitio P3/6K1, de ocurrir, debe tener una tasa de procesamiento muy lenta que no pudo ser detectada en la escala de tiempo utilizada en estas simulaciones.



**Figura 42. Sitio de corte P3/6K1.** A diferencia de los demás sitios específicos por NIa-Pro, esta región presentó una frecuencia de corte del 0%. Probablemente, esto se atribuye a la sustitución por valina en lugar de tirosina en la posición P<sub>3</sub> de la secuencia. La conformación adoptada por la molécula durante 300 ps estuvo asociada a la configuración C (líneas grises).

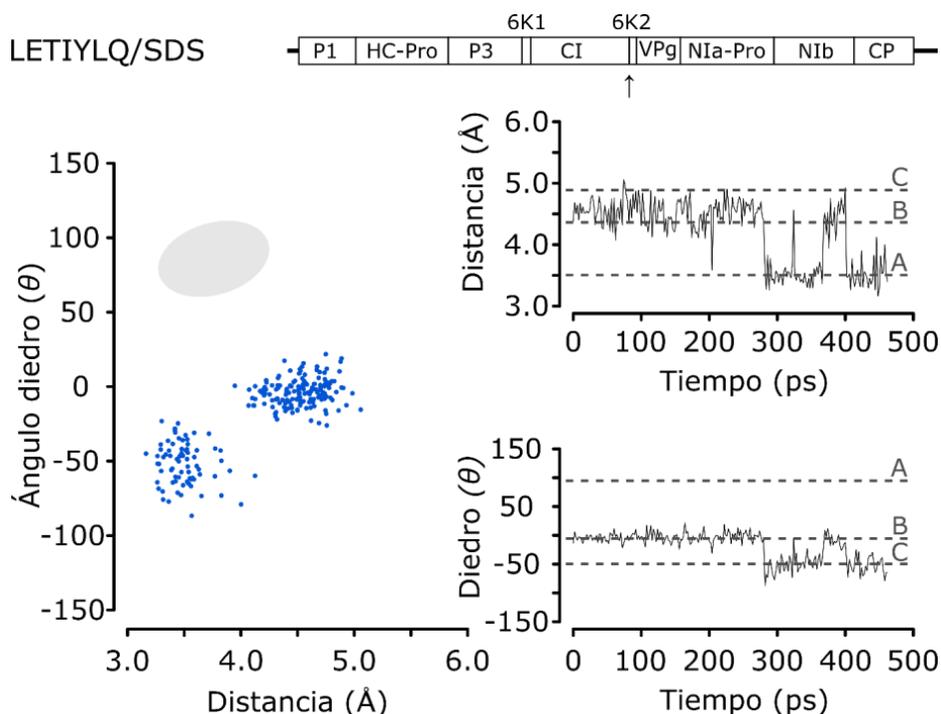
En cuanto al sitio 6K1/CI (Figura 43), se detectó un estado apto para efectuar el corte durante 100 ps, en los cuales el ángulo diedro promedio fue de 85.29°. Sin embargo, aun cuando la distancia azufre-carbono no presentó valores lejanos al promedio 3.85 Å (desviación estándar de 0.25 Å), el 41.38% de los datos estuvo por fuera de la frontera de decisión debido a que el plano peptídico tuvo un ángulo diedro promedio de -45.65°, asemejándose a la configuración C. La secuencia de corte para este intermediario cuenta con residuos conservados en las posiciones críticas para la especificidad de la proteasa por este sustrato, lo cual se ve reflejado en los resultados de la simulación.



**Figura 43. Sitio de corte 6K1/CI.** La frecuencia en la región de corte del 58.62% es coherente con la asociación de la molécula con la configuración A. Sin embargo, durante 160 ps de la simulación, se presentó una orientación del ángulo diedro referente a la configuración C, lo que explica el 41.38% de los datos que estuvieron fuera de la frontera de decisión.

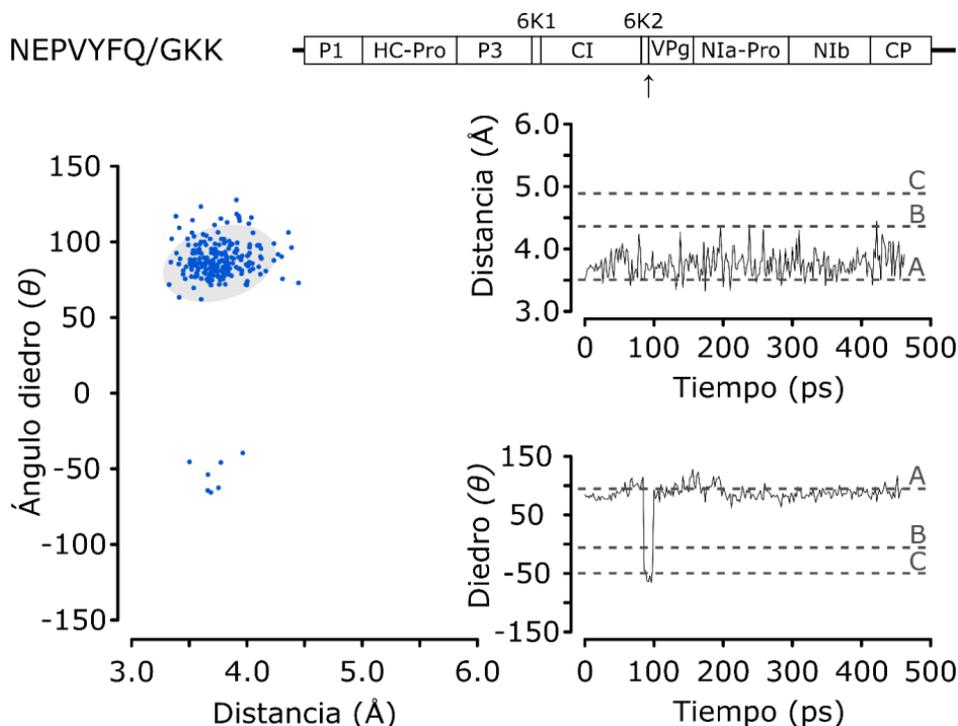
El péptido CI/6K2 siempre estuvo en conformaciones por fuera de la zona de corte. De acuerdo a la distancia azufre-carbono, el péptido-sustrato no presentó una configuración apropiada para la hidrólisis del enlace peptídico. Tampoco lo reportó la orientación entre el plano peptídico y el enlace tiol del ataque nucleofílico. En ambos aspectos, la distancia promedio (4.5 Å) estuvo por arriba de un valor ideal de 3.8 Å (configuración A) y el ángulo diedro (-5.28°) muy por debajo de la orientación buscada (Figura 44). Posterior a los 290 ps, se observaron tres regiones en donde la distancia osciló en un valor de 3.7 Å, pero el ángulo diedro no alcanzó un estado cercano a la configuración A para efectuar el corte, arrojando un valor de frecuencia del 0% y categorizándose en un cluster similar al de la configuración C. Este precursor ha sido reportado en la literatura como uno de los más lentos en procesarse debido a su papel en el proceso replicativo del -RNA (Merits *et al.*, 2002). Al igual que para P3/6K1, se postula que la hidrólisis del

sitio CI/6K2 debe ocurrir a una escala de tiempo superior a la ventana utilizada en esta simulación.



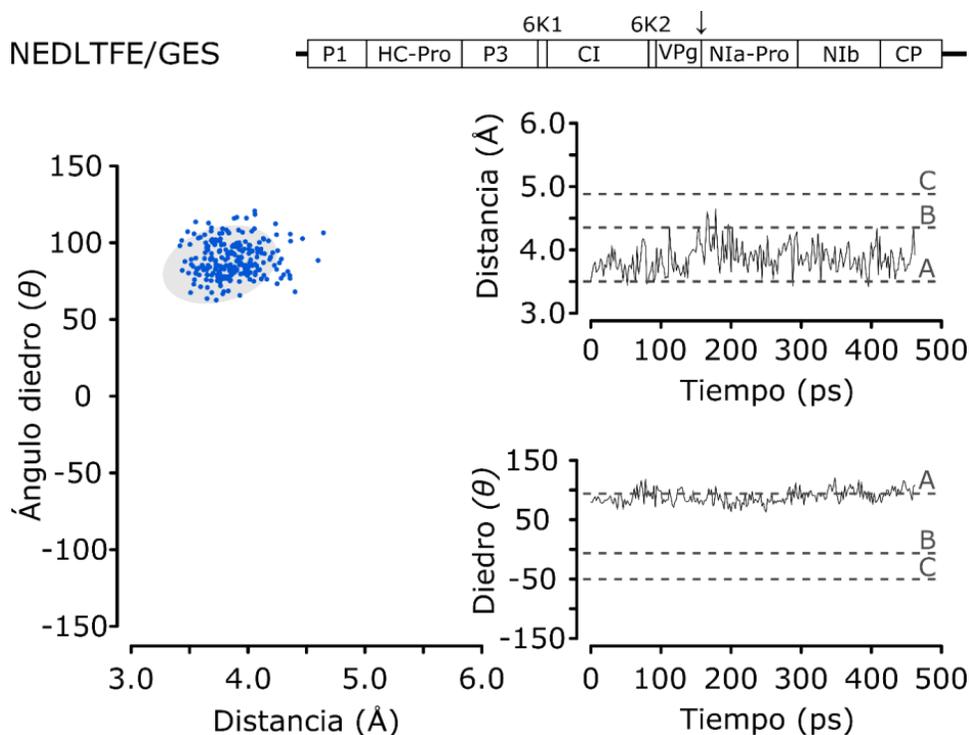
**Figura 44. Sitio de corte CI/6K2.** Si bien esta región es procesada por NIa-Pro, presentó una frecuencia en la región de corte del 0% lo que se ejemplifica por las fluctuaciones entre las configuraciones B y C (líneas grises). En la literatura se reporta que la maduración de este precursor es de las más lentas en efectuarse.

En contraste, la simulación del péptido 6K2/VPg pasó un 86.6% de las veces en la zona de corte (Figura 45) sugiriendo que este precursor es procesado rápidamente por NIa-Pro. Si bien al inicio el ángulo diedro presentó un valor cercano a  $-50^\circ$ , esto solo sucedió por un intervalo de 10 ps, que una vez superado, mantuvo una cifra más apropiada ( $88.8^\circ$ ) por el resto de la simulación. La distancia promedio entre átomos  $S_\gamma$ -C fue de  $3.76 \text{ \AA}$  (desviación estándar de  $0.20 \text{ \AA}$ ), lo cual fue acorde al caso ideal de la configuración A, en el que la misma distancia se estima cercana a este valor. Este resultado es consistente con estudios experimentales donde se ha reportado que la vida media del precursor 6K2/VPg es menor a cinco minutos, lo que indica un alto procesamiento de la enzima (Merits *et al.*, 2002).



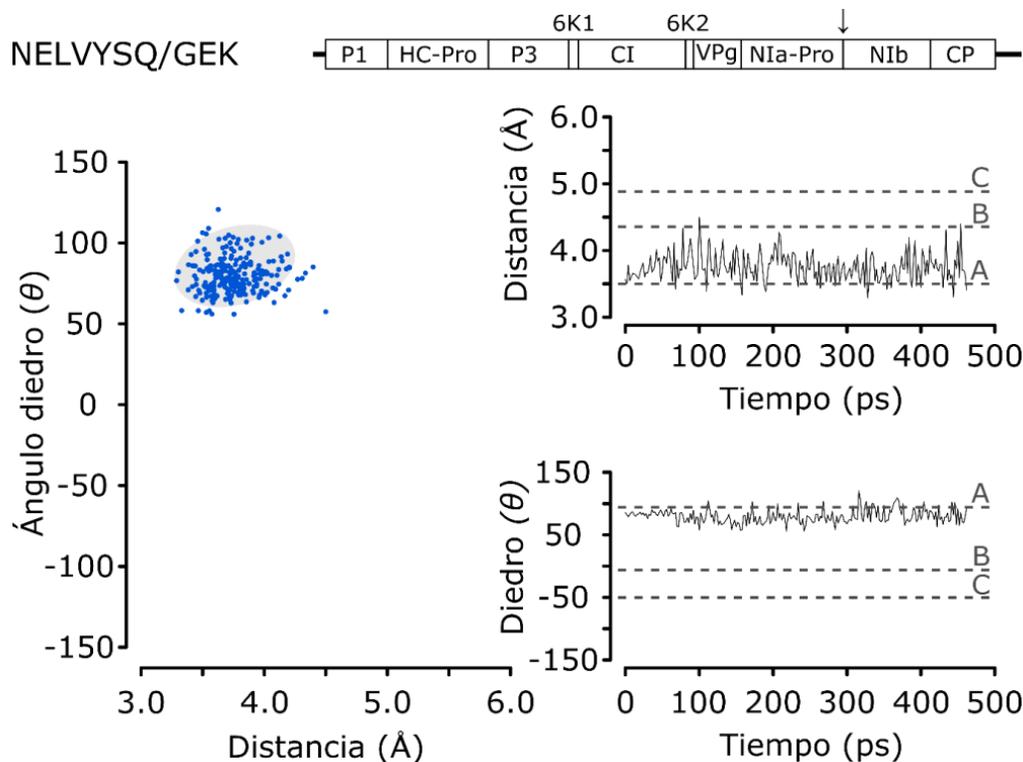
**Figura 45. Sitio de corte 6K2/VPg.** La mayor parte de la simulación este precursor estuvo en la configuración A, presentando una frecuencia en la región de corte del 86.6%.

A diferencia de los demás sitios de corte específicos de NIa-Pro, el precursor VPg/NIa-Pro cuenta con un ácido glutámico en lugar de una glutamina en la posición P<sub>-1</sub> que no parece afectar significativamente el tiempo de permanencia en la región de corte, que fue de 86.6%. La distancia carbono-azufre tuvo un promedio de  $3.87 \pm 0.23$  Å con un valor máximo de 4.6 Å; el promedio del ángulo diedro fue de  $89.02 \pm 11.53^\circ$  (Figura 46).



**Figura 46. Sitio de corte VPg/Nla-Pro.** Al igual que en el sitio 6K2/VPg, esta región presentó un frecuencia en la región de corte del 86.6%. No se observaron fluctuaciones entre los estados B ni C.

En conjunto con la región de los precursores 6K2/VPg, VPg/Nla-Pro y Nlb/CP, el sitio de corte entre los péptidos Nla-Pro y Nlb estuvo entre los mayores valores de frecuencia dentro de la región de corte (90.95%). No se observaron fluctuaciones entre las configuraciones B o C; por el contrario, la distancia entre los átomos  $S_{\gamma}$ -C estuvo en un promedio de 3.75 Å con un valor de ángulo diedro de 79.72° y con desviaciones estándar de 0.21Å y 11.15°, respectivamente; asemejándose más a la configuración A (Figura 47).



**Figura 47. Sitio de corte Nla-Pro/NIb.** Después del sitio NIb/CP, esta región presentó la mayor frecuencia en la región de corte, 90.95%, lo que se evidencia por la prelación del sustrato y la proteína en adoptar la configuración A.

La Tabla 4 resume las frecuencias obtenidas en la región de corte para cada uno de los sitios de la poliproteína, junto con los valores del promedio y la desviación estándar para las variables de distancia y ángulo diedro. Según el modelo propuesto, el sitio NIb/CP presentó la mayor probabilidad de corte en el procesamiento de la poliproteína de TEV, seguido del sitio NIa-Pro/NIb. Las regiones VPg/NIa-Pro y 6K2/VPg obtuvieron el mismo valor de probabilidad para ser procesadas por la proteasa, mientras que el sitio 6K1/CI adquirió la menor de ellas. De acuerdo con las simulaciones moleculares, la región CI/6K2 no presentó un estado en el que el sustrato se asemejara a la configuración A, lo que terminó en una frecuencia de corte del 0% y en una constante cinética de igual magnitud ( $k_3=0$ ).

**Tabla 4. Frecuencia para los sitios de corte simulados en TEV.** En negrita se señalan las posiciones P<sub>-6</sub>, P<sub>-3</sub> y P<sub>-1</sub> relativas al sitio donde ocurre el corte.

Sitio de corte	Secuencia	Frecuencia	Constante cinética	Distancia (Å)		Diedro (°)	
				$\bar{x}$	s	$\bar{x}$	s
P1/HC-Pro	CH <b>SMTHY</b> /SDK	0.2284	—	4.22	0.64	45.00	55.06
HC-Pro/P3	M <b>KTYNVG</b> /GMN	0.0	—	3.76	0.36	-36.05	26.27
P3/6K1	K <b>EDLVEQ</b> /AKQ	0.0	k <sub>1</sub>	4.05	0.41	-37.97	60.34
6K1/CI	R <b>EIIYTQ</b> /SLD	0.5862	k <sub>2</sub>	3.86	0.25	47.65	64.50
CI/6K2	L <b>ETIYLQ</b> /SDS	0.0	k <sub>3</sub>	4.20	0.52	-19.09	24.97
6K2/VPg	N <b>EPVYFQ</b> /GKK	0.8664	k <sub>4</sub>	3.76	0.20	84.44	26.77
VPg/NIa-Pro	N <b>EDLTFE</b> /GES	0.8664	k <sub>5</sub>	3.87	0.23	89.02	11.53
NIa-Pro/NIb	N <b>ELVYSQ</b> /GEK	0.9095	k <sub>6</sub>	3.76	0.21	79.72	11.15
NIb/CP	T <b>ENLYFQ</b> /SGT	0.9397	k <sub>7</sub>	3.76	0.20	85.83	10.32

Este resultado se encuentra respaldado por las investigaciones de Merits *et al.* (2002), quienes mediante experimentos de pulso y caza en PVA, determinaron que los precursores P3-6K1, CI-6K2 y las regiones internas de NIa fueron procesados lentamente por la proteasa; CI-6K2 fue detectado incluso después de 2 horas en la fase del caza. Estos dos eventos de procesamiento fueron evidenciados en las simulaciones moleculares, pues la región P3/6K1 también obtuvo una frecuencia en la región de corte del 0% que puede ser atribuida por la presencia de valina en la posición P<sub>-3</sub> en lugar de tirosina. Este lugar es reportado en la literatura por presentar una elevada conservación entre diversos aislamientos de TEV, por lo que la sustitución por un aminoácido no polar afecta su capacidad de procesamiento (Dougherty *et al.*, 1988).

Es importante aclarar que los valores del 0% no indican que la hidrólisis peptídica no se lleve a cabo en estos sitios, sino que sus respectivas frecuencias de corte presentaron un valor inferior al margen de detección del modelo:  $1/232 = 0.0043$  y no pudieron ser detectadas. Lo anterior sugiere que estas regiones requieren una ventana mayor del tiempo de simulación para obtener una estimación más precisa de su probabilidad de corte.

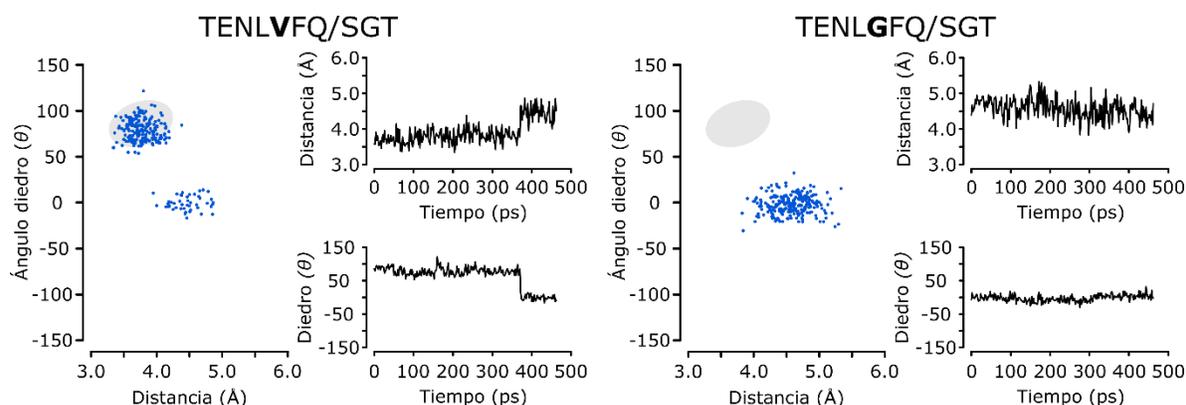
Finalmente, el modelo también predice que el sitio P1/HC-Pro, cuyo procesamiento es llevado a cabo por la proteasa P1, podría ser procesado por

Nla-Pro al encontrarse que este péptido-sustrato permanece 22.84% de las veces en la configuración A durante 200 ps de simulación. La validez de esta predicción puede ser verificada experimentalmente; sin embargo, este resultado carece de sentido biológico puesto que se espera que el sitio P1/HC-Pro sea procesado autocatalíticamente en la fase temprana de la síntesis de la poliproteína, cuando aún no está disponible Nla-Pro (Ivanov *et al.*, 2014).

#### **4.1.7.1. Análisis de mutaciones en el sitio de corte Nlb/CP**

Dougherty y colaboradores realizaron un estudio mutacional sobre la tolerancia que tiene la proteasa frente a diversos aminoácidos en el sitio de corte. Entre los resultados de su trabajo se encuentra el reporte del motivo E-X-X-Y-X-Q/(SG) como blanco de corte de Nla-Pro en TEV que permitió concluir que el cambio de aminoácidos en estas posiciones afecta la efectividad de la enzima. Con el fin de darle soporte experimental al modelo, se realizaron simulaciones con mutantes en la posiciones P<sub>-3</sub> del sitio de corte Nlb/CP donde se sustituye tirosina por valina o glicina. Según Dougherty *et al.* (1988), la enzima con valina en la posición P<sub>-3</sub> presentó una tasa de corte mucho menor a la proteasa natural de TEV, mientras que la segunda (glicina en lugar de tirosina) abolió por completo su capacidad proteolítica. Los resultados de la simulación se muestran en la Figura 48. En la sustitución por valina se observó una frecuencia en la región de corte del 69.83%, lo que significa que hubo una reducción del 24.14% respecto al sitio original. En su estudio, Dougherty *et al.*, dejan claro que el precursor Nlb/CP puede ser procesado en un punto subóptimo cuando se preserva la naturaleza hidrofóbica de las posiciones P<sub>-5</sub>, P<sub>-4</sub> y P<sub>-2</sub>. No obstante, al ser reemplazada la posición P<sub>-3</sub> siempre se observa una disminución en la función del sustrato. El resultado de la simulación es acorde con esta conclusión puesto que el péptido-sustrato presenta dos estados diferentes, pero solo el de la zona limitada por la frontera de decisión es más factible de ser hidrolizado. La configuración minoritaria (30.17%), sucedida a partir de los 380 ps, hace que esta secuencia no esté en un escenario óptimamente procesable, pues la distancia S<sub>γ</sub>-C promedio pasa de 3.8 Å a 4.46 Å y el ángulo diedro promedio es de -165°, en contraste con una media inicial de

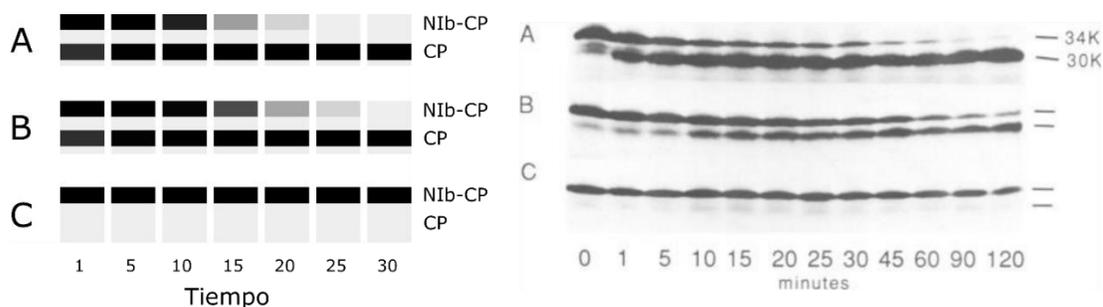
76.61°, en los primeros 300 ps de la simulación; por lo tanto, la enzima y el sustrato adoptan una conformación similar a la configuración B. Los resultados para la sustitución por glicina también fueron acordes con el trabajo experimental de Dougherty *et al.* Para esta simulación se obtuvo una frecuencia de 0% en la región de corte, con una distancia promedio azufre-carbono de 4.57 Å y un valor de ángulo diedro promedio de -2.34°. La sustitución de un aminoácido polar neutro por residuos de naturaleza hidrofóbica en P<sub>-3</sub> tiene una influencia directa en la interacción proteasa-sustrato, al término que es posible abolir la capacidad de hidrolizar el enlace peptídico.



**Figura 48. Simulación de las mutaciones en el sitio de corte Nib/CP.** La simulación para la sustitución por valina en la posición P<sub>-3</sub> (izquierda) muestra un punto subóptimo de procesamiento proteolítico al evidenciarse dos escenarios que pueden ser adoptados por el sustrato en presencia del sitio activo de la proteasa. En la mutación por glicina (derecha) no se observó, durante los 460 ps de la simulación, ninguna configuración apropiada para ser procesada por la proteasa.

La Figura 49 muestra un paralelo entre los resultados de las mutaciones simuladas y los experimentos reales del trabajo de Dougherty y colaboradores. Con el propósito de simular un gel de electroforesis que representa la tasa de desaparición del precursor Nib-CP, a lo largo de una escala relativa de tiempo, se utilizó el modelo de maduración descrito empleando las constantes cinéticas reportadas en la Tabla 4. Como puede verse, el modelo presentado da una buena descripción cualitativa de la eficiencia de corte y permite hacer predicciones que

pueden ser verificadas experimentalmente mediante técnicas convencionales de biología molecular.



**Figura 49.** Tiempo del procesamiento de la poliproteína Nib-CP. En el panel izquierdo se ilustra la simulación de un gel de electroforesis según las constantes cinéticas halladas en el modelo de maduración propuesto. El segmento A representa la proteólisis del péptido llevada a cabo por una N1a-Pro de TEV sin mutaciones en la secuencia de corte. El segmento B corresponde a la simulación de la enzima con la sustitución de tirosina por valina en la posición P<sub>-3</sub>. En este caso el procesamiento ocurre a una tasa mucho menor que en el escenario anterior. El segmento C presenta la mutación con glicina que anula completamente la capacidad catalítica de la enzima pues no logra generarse CP. Los resultados son comparados con un gel de electroforesis real realizado en el trabajo de Dougherty *et al.* (panel derecho). Las unidades de tiempo del gel simulado (izquierda) son relativas a una escala arbitraria y no corresponden a las presentadas en gel de la derecha.

#### 4.1.8. Modelo de maduración cinética en el sistema PVY

El modelo de maduración enzimática fue también evaluado en dos genomas de PVY identificados en el municipio de La Unión (Antioquia) que fueron obtenidos a partir de NGS. La metodología y los resultados de la investigación se presentan en formato de artículo en el capítulo 4.3 de este trabajo.

Los aislamientos de los genomas se agruparon en dos clados diferentes dentro del grupo PVY<sup>N/NTN</sup>. El primero de ellos corresponde al genoma de PVY procedente de tejido foliar (f\_PVY), el cual se reportó asociado al grupo de cepas recombinantes PVY<sup>NTN</sup> y que además presentó gran similitud con el genoma de PVY secuenciado en el municipio de Yarumal (reportado en el capítulo 4.2 de este trabajo), sugiriendo un origen común entre ellos. Por otro lado, el segundo aislamiento de PVY se reportó infectando tubérculo-semilla (t\_PVY) con una afinidad filogenética por las cepas del grupo necrosante de venas en tabaco,

PVY<sup>N</sup>. Los síntomas causados por los virus de este grupo no se consideran severos en papa, a diferencia del aislamiento recombinante PVY<sup>NTN</sup>, fuertemente asociado con la aparición de anillos necróticos en tubérculos de papa; enfermedad conocida como PTNRD (*Potato Tuber Necrotic Ringspot Disease*).

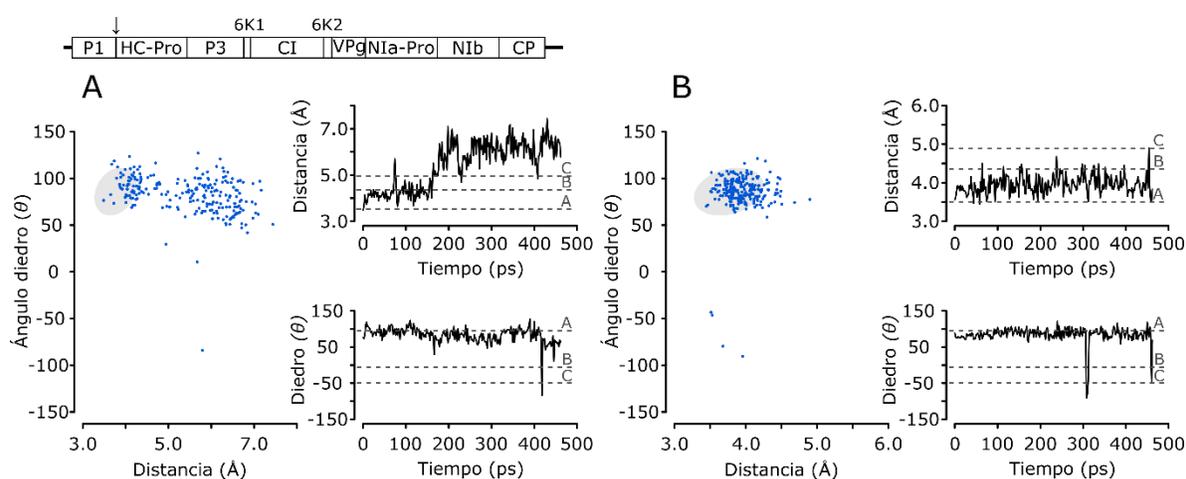
Los resultados de la simulación molecular en el sistema PVY fueron discriminados por la frontera de decisión descrita para TEV (Figura 38). Todos los sitios de corte de ambos genomas presentaron una conservación en la secuencia, excepto por las regiones P1/HC-Pro, CI/6K2 y 6K2/VPg. Sin embargo, las diferencias en la secuencia para estos dos últimos sitios no comprometieron posiciones en el motivo de reconocimiento de la proteasa, pues se encuentran ubicadas en la posición P<sub>+2</sub> (Figura 50). En la región P1/HC-Pro las diferencias entre aminoácidos se ubican en la posición P<sub>-7</sub> (antes del sitio de corte); y en P<sub>+1</sub>, P<sub>+2</sub> y P<sub>+3</sub>, después del corte. Estas mutaciones no se consideran significativas en el planteamiento del modelo, pues hacen parte del sitio autocatalítico independiente del mecanismo de maduración de NIa-Pro.

<b>P1/HC-Pro</b>	P <sub>-6</sub> P <sub>-3</sub> P <sub>-1</sub>	<b>6K1/CI</b>	P <sub>-6</sub> P <sub>-3</sub> P <sub>-1</sub>	<b>VPg/NIa-Pro</b>	P <sub>-6</sub> P <sub>-3</sub> P <sub>-1</sub>
f_PVY	MDSMVQFGVM	f_PVY	DHEVRHQSLD	f_PVY	KQEVEHEAKS
t_PVY	L . . . . . SAE	t_PVY	. . . . .	t_PVY	. . . . .
<b>HC-Pro/P3</b>		<b>CI/6K2</b>		<b>NIa-Pro/Nib</b>	
f_PVY	IKHYRVGGIP	f_PVY	LQFVHHQAAT	f_PVY	HDVVVEQAKH
t_PVY	. . . . .	t_PVY	. . . . . T .	t_PVY	. . . . .
<b>P3/6K1</b>		<b>6K2/VPg</b>		<b>Nib/CP</b>	
f_PVY	EYDVRHQIRST	f_PVY	VETVSHQGN	f_PVY	TYEVHHQGND
t_PVY	. . . . .	t_PVY	. . . . . R .	t_PVY	. . . . .

**Figura 50. Análisis de secuencias para los sitios de corte de f\_PVY y t\_PVY.** Ambos genomas de PVY reportaron secuencias conservadas para la mayoría de los sitios de corte. Las únicas excepciones se encontraron en las posiciones P<sub>+2</sub> de las regiones CI/6K2 y 6K2/VPg. Adicionalmente, el sitio autocatalítico P1-HC/Pro presentó diferencias en los residuos en P<sub>-7</sub>, P<sub>+1</sub>, P<sub>+2</sub>, P<sub>+3</sub>.

El sitio de corte P1/HC-Pro simulado para f\_PVY obtuvo una distancia S<sub>γ</sub>-C promedio de 5.42 Å ± 1.01 Å, dos unidades alejada de la distancia óptima de la configuración A. El valor del ángulo diedro promedio fue de 83.05° ± 20.59°. Como se observa en la Figura 51-A, durante los primeros 100 ps de la simulación, la

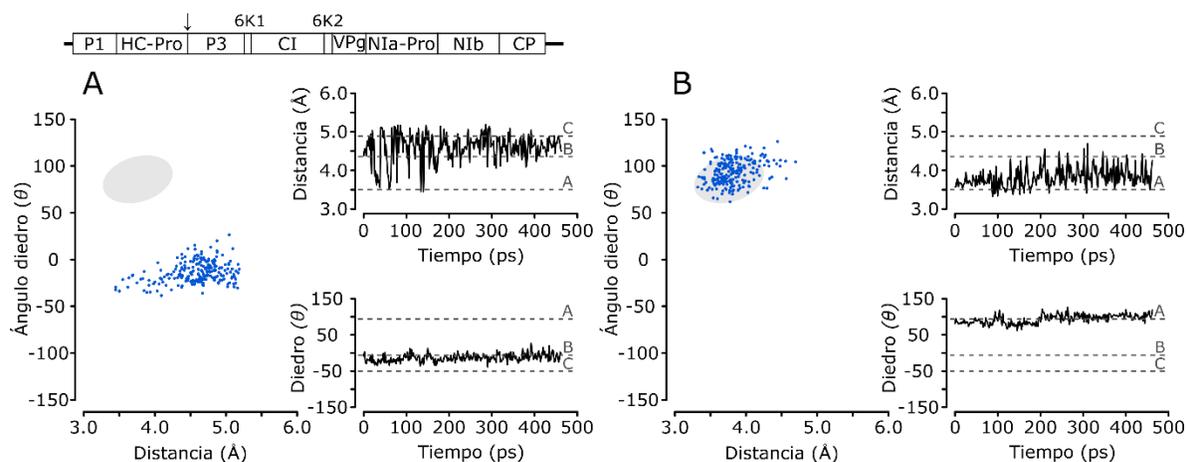
conformación del sustrato y la proteasa estuvo relacionada con la configuración A, por lo que algunos puntos cayeron dentro de la región de corte, aportando una frecuencia del 15.95%. Similarmente en  $t_{PVY}$  se reportó un ángulo diedro promedio de  $84.47^\circ \pm 22.82^\circ$  y una distancia promedio entre el azufre y el carbono carbonílico de  $3.95 \text{ \AA} \pm 0.24 \text{ \AA}$ ; valores muy relacionados con la configuración A por lo que la frecuencia de corte fue del 81.03% (Figura 51-B). Si bien el modelo indica que este sitio puede ser procesado con inclusive altas constantes cinéticas relativas, no es viable biológicamente por ser regiones probablemente procesadas antes de generarse NIa-Pro (Ivanov *et al.*, 2014).



**Figura 51. Sitio de corte P1/HC-Pro del sistema PVY. A)** La frecuencia de corte para el primer sitio autocatalítico en  $f_{PVY}$  fue del 15.95%. **B)** En  $t_{PVY}$  la frecuencia de corte fue del 81.03%. Al igual que en TEV, las simulaciones moleculares indican una probabilidad de corte para este sitio. Sin embargo, esto puede ser refutado experimentalmente pues se sabe que P1 cataliza la separación de la poliproteína en su extremo C-terminal.

El sitio HC-Pro/P3 presentó resultados contradictorios entre ambos aislamientos. En  $f_{PVY}$  se obtuvo una distancia promedio  $S_Y-C$  de  $4.59 \text{ \AA} \pm 0.36 \text{ \AA}$ , lo cual corresponde a valores observados entre las configuraciones B y C; sin embargo, durante los primeros 100 ps de la simulación, esta distancia fluctuó entre la configuración A. El ángulo diedro promedio fue de  $-16.27^\circ \pm 11.14^\circ$ , por lo que en general la molécula tuvo un comportamiento más cercano a la configuración B. Como se observa en la Figura 52-A, la frecuencia en la región de corte fue del 0%. Contrario a este resultado, en  $t_{PVY}$  se obtuvo una distancia  $S_Y-C$  de  $3.82 \text{ \AA} \pm$

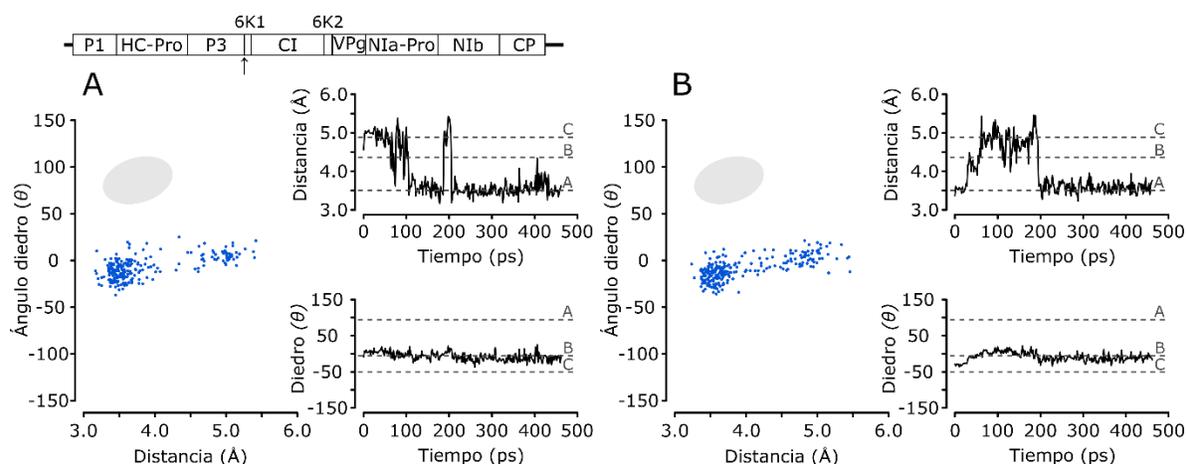
0.25 Å y un ángulo diedro promedio de  $93.64^\circ \pm 12.40^\circ$ , valores correspondientes a la configuración A, y una frecuencia de corte del 66.38%. Lo anterior difiere notoriamente de los respectivos escenarios en los sistemas TEV y f\_PVY, inclusive cuando este último presenta la misma secuencia del péptido-sustrato simulado (Figura 52-B). Si bien el modelo permite clasificar este sitio en t\_PVY como apto para realizar el corte, no tiene sentido en la biología del virus puesto que HC-Pro corta su propio extremo C-terminal para separarse de la poliproteína (Valli *et al.*, 2014).



**Figura 52. Sitio de corte HC-Pro/P3 del sistema PVY. A)** La frecuencia de corte para el segundo sitio autocatalítico en f\_PVY fue del 0%, a diferencia del **B)** 66.38% obtenido para la misma secuencia del sustrato en t\_PVY. Este sitio es independiente de la actividad proteolítica de NIa-Pro.

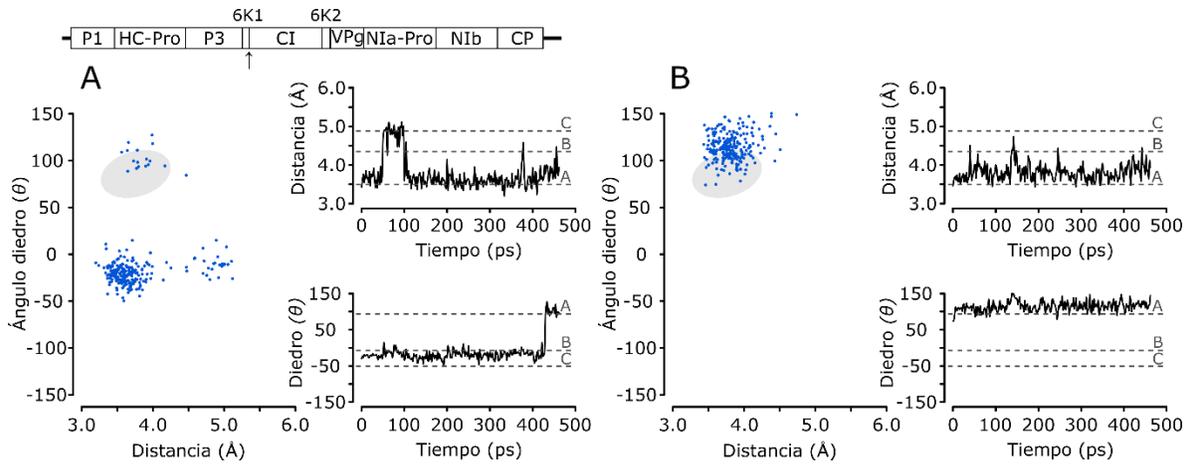
Como se puede observar en la Figura 53-A, el comportamiento de la molécula en las simulaciones del sitio P3/6K1 de f\_PVY estuvo más asociado a la configuración B, con fluctuaciones hacia la configuración C ocurridas en la distancia S<sub>Y</sub>-C durante los primeros 100 ps. El valor promedio de la distancia fue del  $3.88 \text{ \AA} \pm 0.62 \text{ \AA}$ ; mientras que para el ángulo diedro se obtuvo un promedio de  $-7.24^\circ \pm 11.72^\circ$ , también relacionado con la configuración B y sin la ocurrencia de saltos durante los 460 ps totales. Estos valores en conjunto aportaron una frecuencia de corte del 0%. En cuanto a t\_PVY, el sitio de corte P3/6K1 presentó un valor promedio de distancia S<sub>Y</sub>-C de  $3.98 \text{ \AA} \pm 0.59 \text{ \AA}$  y ángulo diedro de  $-8.78^\circ \pm 12.12^\circ$ . Durante aproximadamente 240 ps del final de la simulación, el complejo

adoptó una conformación cercana a la configuración A, exclusivamente en términos de la distancia carbono carbonílico y azufre. El resto del tiempo presentó saltos entre las configuraciones B y C y por tanto su frecuencia en la región de corte fue también del 0% (Figura 53-B). Al igual que lo ocurrido en el sistema de TEV, las simulaciones sugieren que este sitio de corte presenta una tasa de procesamiento lenta y fuera del margen de detección del modelo planteado, por lo que se requeriría un tiempo de simulación mayor para determinar este parámetro.



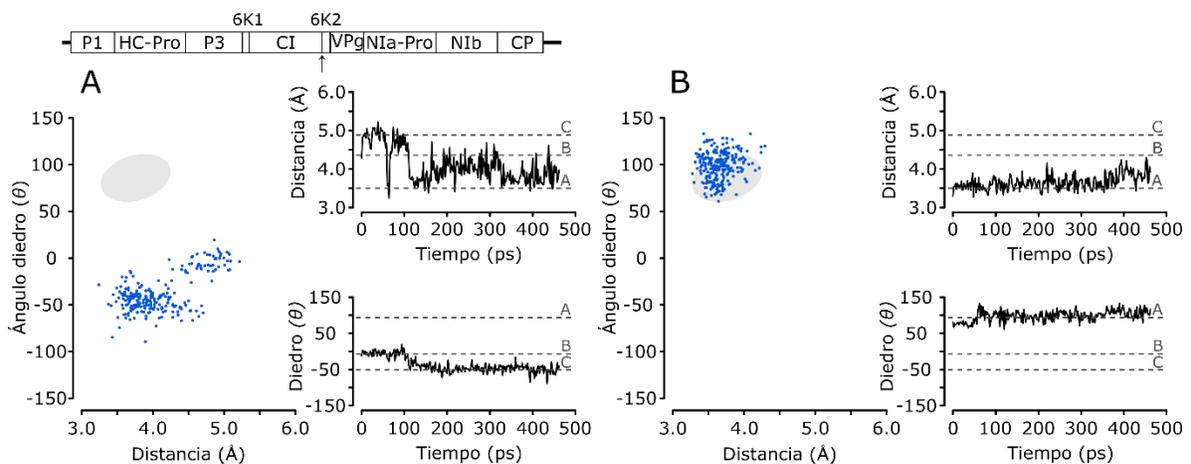
**Figura 53. Sitio de corte P3/6K1 del sistema PVY.** Tanto para ambos genomas procedentes de **A)** tejido foliar y **B)** tubérculo se obtuvieron frecuencias del 0% en la región de corte. Los resultados difieren de la simulación realizada en el correspondiente sitio de TEV.

El sitio 6K1/CI de f\_PVY obtuvo una distancia  $S_{\gamma-C}$  promedio de  $3.78 \text{ \AA} \pm 0.43 \text{ \AA}$ , relacionada con la configuración A, y un ángulo diedro promedio de  $-10.88^\circ \pm 30.70^\circ$ , propio de la configuración B (Figura 54-A). La frecuencia de corte fue del 4.31%. En contraste, en t\_PVY la frecuencia obtenida fue del 33.62% con valores de  $3.80 \text{ \AA} \pm 0.22 \text{ \AA}$  y  $114.59^\circ \pm 14.22^\circ$ , para la distancia azufre-carbono carbonílico y ángulo diedro, respectivamente. Como se ilustra en la Figura 54-B, la orientación del ángulo diedro estuvo favorecida hacia la configuración A, mientras que la distancia entre los átomos osciló entre las configuraciones A y B.



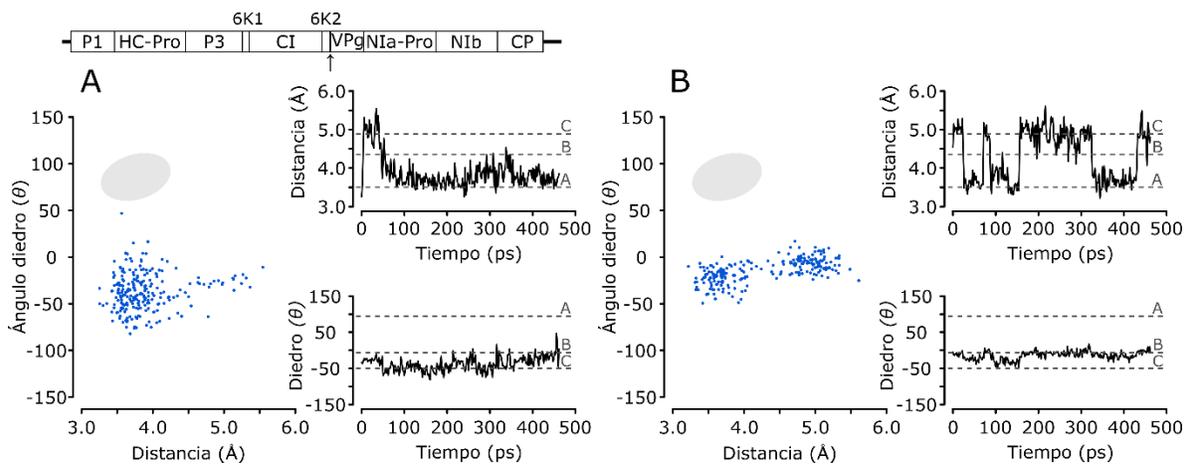
**Figura 54. Sitio de corte 6K1/CI del sistema PVY. A)** La frecuencia de corte de esta región en  $f_{PVY}$  fue del 4.31% que difiere significativamente de la obtenida en **B)**  $t_{PVY}$ , la cual fue del 33.62%

La región CI/6K2 aportó el primer dato de frecuencia de corte para  $f_{PVY}$ , correspondiente a un 4.31%. Como se ilustra en la Figura 54-A, la orientación de la proteasa y el sustrato tuvo prevalencia por la configuración C debido a un valor de ángulo diedro promedio de  $-37.22^\circ \pm 20.67^\circ$ . Por el contrario, la distancia  $S_Y-C$  promedio fue del  $4.10 \text{ \AA} \pm 0.45 \text{ \AA}$ , que a partir de los 100 ps presentó valores correspondientes a la configuración A. En  $t_{PVY}$ , se obtuvo una distancia y ángulo diedro promedio de  $3.66 \text{ \AA} \pm 0.19 \text{ \AA}$  y  $98.39^\circ \pm 15.13^\circ$ , respectivamente. Estas cifras contribuyeron a una frecuencia de corte del 66.81%; resultado que difiere del lento procesamiento obtenido en el respectivo sitio de TEV.



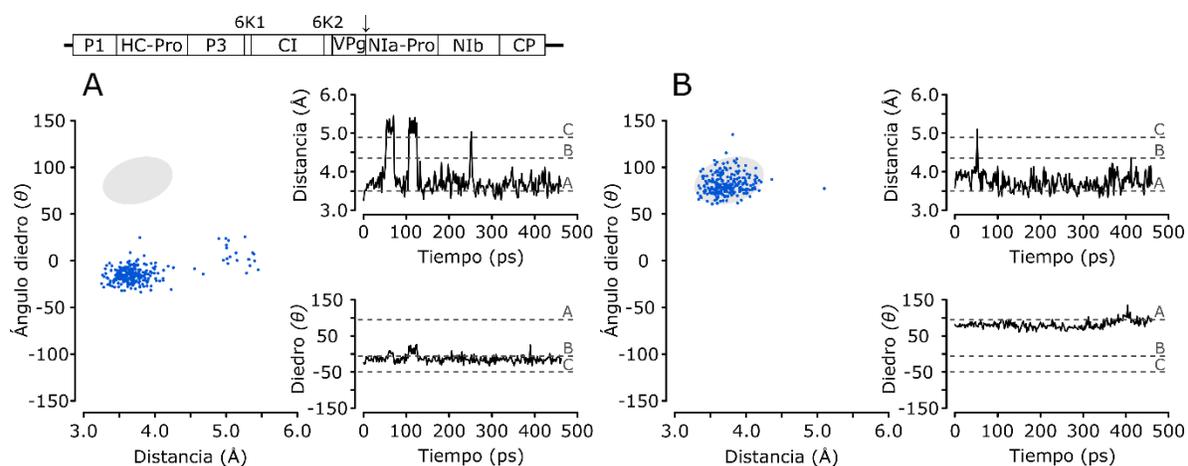
**Figura 55. Sitio de corte CI/6K2 del sistema PVY.** **A)** En  $f_{PVY}$ , la frecuencia del 0% obtenida es coherente con el bajo procesamiento del correspondiente sitio en TEV. Sin embargo, en **B)**  $t_{PVY}$ , tal valor fue del 66.81%.

En contraste con TEV, las simulaciones del péptido 6K2/VPg para ambos genomas de PVY reportaron frecuencias de corte del 0%. Si bien en  $f_{PVY}$  la orientación entre la proteasa y el sustrato estuvo asociada con la configuración A, en términos de la distancia S-C, el valor del ángulo diedro no permitió un acercamiento a la frontera de decisión (Figura 56-A). Los promedios de estas dos variables fueron de  $3.89 \text{ \AA}$  con una desviación estándar de  $0.47 \text{ \AA}$  y  $-37.31^\circ \pm 19.71^\circ$ , respectivamente. Similarmente,  $t_{PVY}$  obtuvo un diedro promedio de  $-14.29^\circ \pm 11.98^\circ$  que, junto a las diversas fluctuaciones en la distancia azufre-carbono carbonílico, permiten relacionarlo con la configuración C (Figura 56-B). Estos resultados difieren además del reporte de Merits y colaboradores (2002), quienes obtuvieron un rápido procesamiento para esta región (menor a 5 min) en PVA. El modelo postulado requiere de un mayor tiempo de simulación, por lo menos en el sistema de PVY planteado, para lograr apreciar la magnitud de estas frecuencias.



**Figura 56. Sitio de corte 6K2/VPg del sistema PVY.** Las frecuencias de corte para ambos sitios en la poliproteína de **A)**  $f_{PVY}$  y **B)**  $t_{PVY}$  fueron del 0%. En TEV este dato fue del 86.64%, por lo que probablemente la diferencia en secuencias en la región de reconocimiento este implicada en el desempeño de las simulaciones moleculares para esta región de corte.

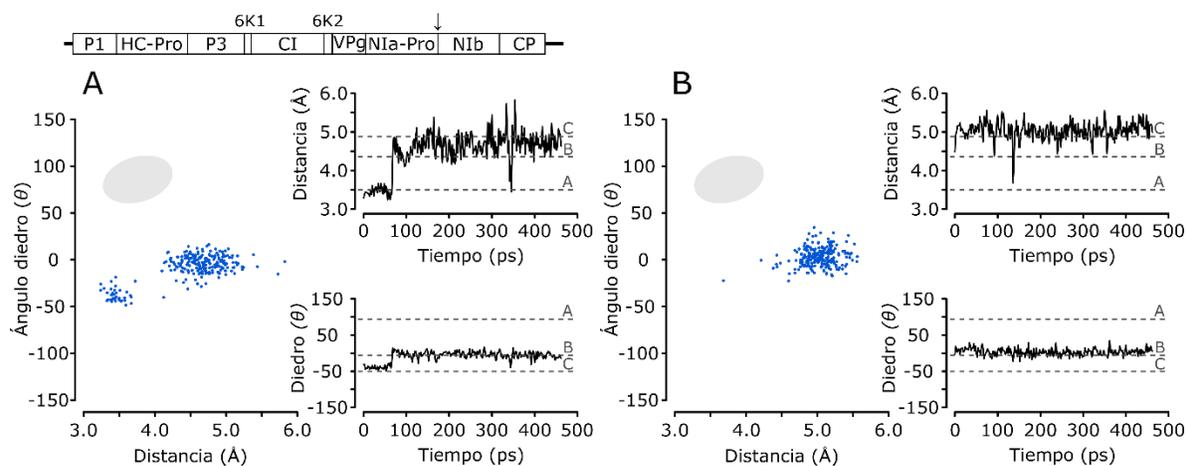
El sitio de corte VPg/N1a-Pro de t\_PVY presentó una frecuencia del 94.4%. El valor promedio del ángulo diedro fue de  $81.80^\circ \pm 10.74^\circ$  y la distancia S<sub>γ</sub>-C de  $3.73 \text{ \AA} \pm 0.23 \text{ \AA}$ , ambos valores acordes con la configuración A (Figura 57-B). La presencia del ácido glutámico en lugar de glutamina en la posición P-1 de este precursor tampoco mostró afectar la afinidad del sustrato y la proteasa por adoptar una orientación óptima para el procesamiento. Sin embargo, para f\_PVY no se logró apreciar una frecuencia de corte, por lo que esta puede ser inferior al 0.43% y será necesario un mayor tiempo de simulación. El promedio obtenido para la distancia fue de  $3.81 \text{ \AA} \pm 0.47 \text{ \AA}$ ; la cual presentó fluctuaciones durante los 100 primeros picosegundos de la simulación. El ángulo diedro promedio fue de  $-13.47^\circ \pm 10.38^\circ$ , muy cercano a la configuración B y parcialmente a la C.



**Figura 57. Sitio de corte VPg/N1a-Pro del sistema PVY. A)** La frecuencia en la región de corte fue del 0% en f\_PVY comparado con **B)** un 94.40% en t\_PVY.

Tampoco se detectó una frecuencia de corte en el precursor N1a-Pro/CP de ambos genomas virales. Para f\_PVY tanto la distancia como el ángulo diedro presentaron valores propios de la configuración B:  $4.50 \text{ \AA} \pm 0.52 \text{ \AA}$  y  $-9.18^\circ \pm 14.95^\circ$ , respectivamente. Solo al inicio de la simulación (primeros 50 ps) se adquirió una distancia óptima de la configuración A, pero debido al desfase entre los planos de interacción nucleofílica, esta conformación no trascendió por el resto del tiempo establecido (Figura 58-A). En t\_PVY se observó que la mayoría de los

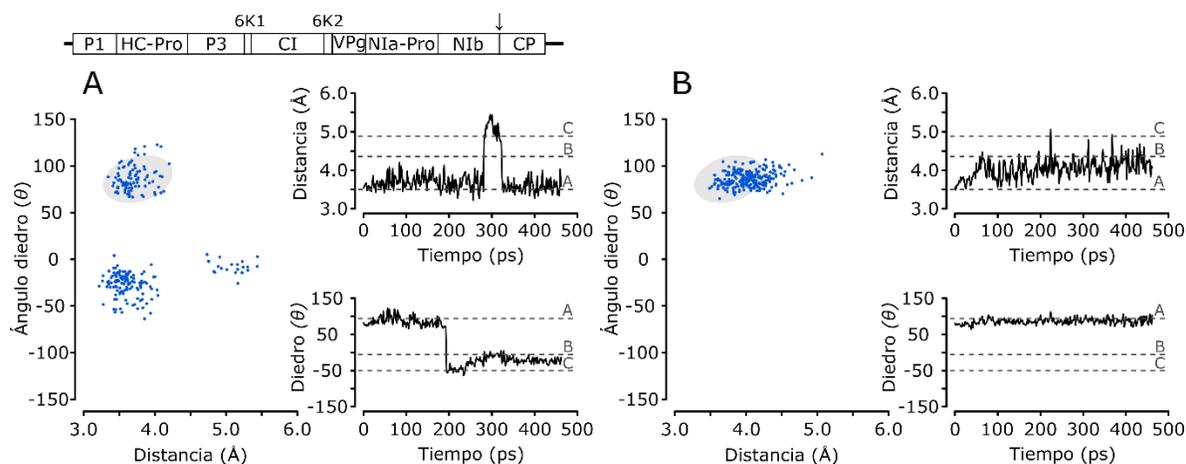
puntos de la simulación se concentraron en una región característica de la configuración B. El ángulo diedro promedio fue de  $3.24^\circ \pm 9.91^\circ$  y la distancia  $S_V-C$  promedio de  $5.03 \text{ \AA} \pm 0.24 \text{ \AA}$ . En contraste con lo obtenido en TEV, los resultados indican que este intermediario presenta una tasa de procesamiento mucho más lenta, lo cual puede afectar las reacciones que ocurran de manera intermolecular, pues de acuerdo con el modelo postulado la proteasa Nla-Pro presenta una baja probabilidad de corte para separarse de la poliproteína.



**Figura 58. Sitio de corte Nla-Pro/Nlb del sistema PVY.** Ambos casos de los aislamientos de **A)** foliar y **B)** tubérculo de PVY presentaron una frecuencia en esta región de corte del 0%.

Al igual que en el sistema de TEV, el sitio que reportó la mayor frecuencia de corte según las simulaciones moleculares fue Nlb/CP, con valores del 36.64% y 73.91% para  $f_{PVY}$  y  $t_{PVY}$ , respectivamente. En la simulación de  $f_{PVY}$  se obtuvieron dos momentos de discriminación entre diferentes configuraciones de corte. Durante los primeros 200 ps se observó un estado similar a la configuración A, pues la distancia promedio entre los átomos  $S_V-C$  fue de  $3.6 \text{ \AA}$  y el ángulo diedro promedio de  $80.9^\circ$ . Posteriormente, y durante 52 ps, ocurrió un sobresalto en la distancia entre átomos  $S_V-C$  a un valor promedio de  $4.75 \text{ \AA}$  y un ángulo de  $-11.07^\circ$ , valores muy extremos propios de la configuración C. La fluctuación entre estos dos estados determinó la frecuencia más pequeña para este precursor entre las tres proteasas simuladas (Figura 59-A). En el caso de  $t_{PVY}$ , el valor del

ángulo diedro promedio ( $86.48^\circ \pm 8.47^\circ$ ) no presentó fluctuaciones durante los 460 ps de la simulación, sugiriendo una orientación adecuada entre los planos donde ocurre la interacción nucleofílica. La mayor diferencia respecto a las distancias de referencia de la configuración A se observó en la longitud promedio del carbono carbonílico y el azufre que fue de  $4.04 \text{ \AA} \pm 0.27 \text{ \AA}$  (Figura 59-B).



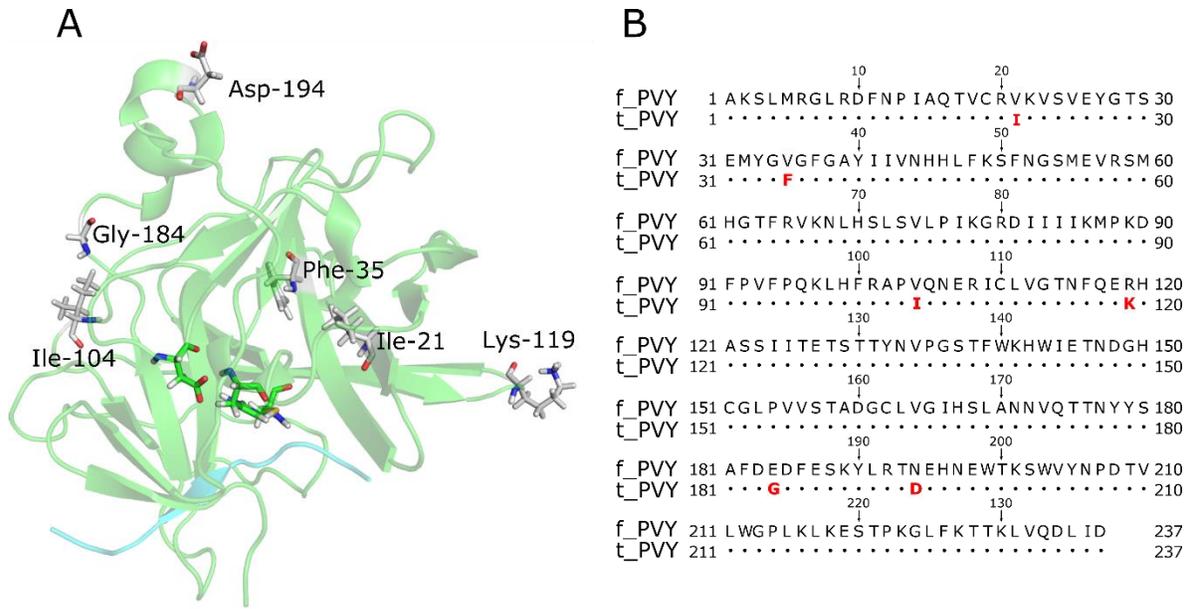
**Figura 59. Paralelo entre las regiones de corte N1b/CP en el sistema PVY. A)** Resultados de la simulación para f\_PVY. La frecuencia en la región de corte fue del 36.64%, siendo la más baja entre los tres virus simulados para este péptido-sustrato. **B)** La frecuencia en la región de corte para t\_PVY fue del 73.91%.

La Tabla 5 resume las frecuencias obtenidas en la región de corte para cada uno de los sitios de la poliproteína, junto con los valores del promedio y la desviación estándar para las variables de distancia y ángulo diedro.

**Tabla 5.** Resultados de las frecuencias de corte y mediciones de distancia  $S_V-C$  y ángulo diedro obtenidas en el sistema PVY.

Sitio de corte	Frecuencia de corte	f_PVY				t_PVY				
		Distancia (Å)		Diedro (°)		Frecuencia de corte	Distancia (Å)		Diedro (°)	
		$\bar{x}$	s	$\bar{x}$	s		$\bar{x}$	s	$\bar{x}$	s
P1/HC-Pro	0.1595	5.42	1.01	83.05	20.59	0.8103	3.95	0.24	84.47	22.82
HC-Pro/P3	0.0	4.59	0.36	-13.33	11.14	0.8276	3.82	0.25	93.64	12.40
P3/6K1	$k_1=0.0$	3.88	0.62	-7.24	11.71	$k_1=0.0$	3.98	0.59	-8.78	12.12
6K1/CI	$k_2=0.0431$	3.78	0.43	-10.88	30.70	$k_2=0.3362$	3.80	0.22	114.59	14.22
CI/6K2	$k_3=0.0$	4.10	0.45	-37.22	20.67	$k_3=0.6681$	3.66	0.19	98.39	15.13
6K2/VPg	$k_4=0.0$	3.89	0.42	-37.31	19.71	$k_4=0.0$	4.32	0.66	-14.29	11.98
VPg/N1a-Pro	$k_5=0.0$	3.81	0.47	-13.47	10.38	$k_5=0.9440$	3.73	0.23	81.80	10.74
N1a-Pro/N1b	$k_6=0.0$	4.50	0.52	-9.18	14.95	$k_6=0.0$	5.03	0.24	3.24	9.91
N1b/CP	$k_7=0.3664$	3.78	0.46	21.61	56.81	$k_7=0.7391$	4.04	0.27	86.48	8.47

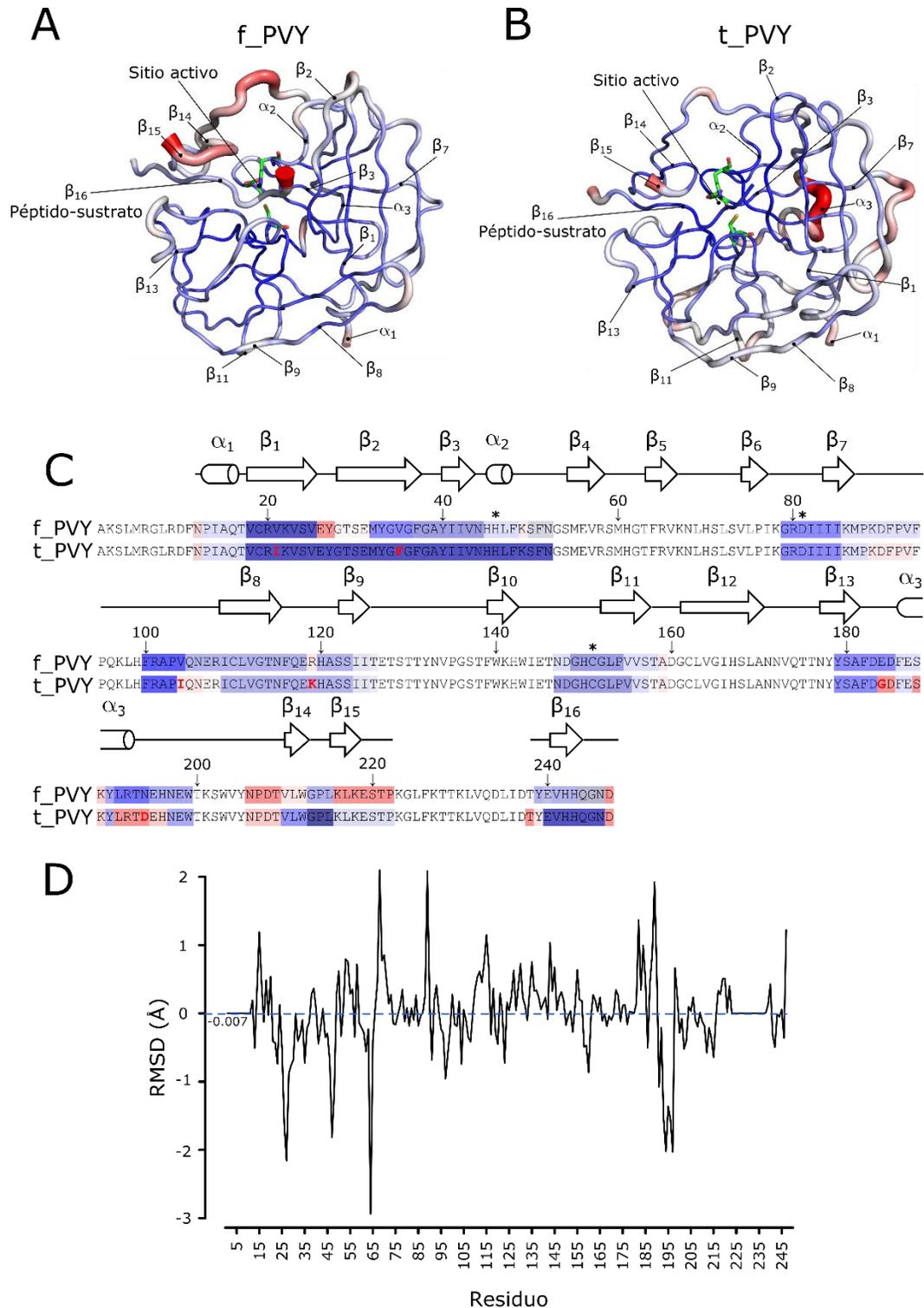
Los resultados de las simulaciones moleculares permiten concluir que la tasa de procesamiento de la poliproteína en el sistema PVY difiere significativamente entre ambos aislamientos. Sin embargo, resulta importante discutir las discrepancias entre los valores de las constantes cinéticas halladas para una misma secuencia en la región de corte simulada. Probablemente, el contraste en los sitios CI/6K2 y VPg/Nla-Pro esté asociado con la secuencia de la proteasa Nla-Pro de sus respectivos genomas. Se sabe que PVY<sup>N</sup> (t\_PVY) y PVY<sup>NTN</sup> (f\_PVY), aun siendo cepas de naturaleza no recombinante y recombinante, respectivamente, comparten una misma región de aproximadamente 3 kb entre el dominio VPg y parte de su cápside (Lorenzen *et al.*, 2006). No obstante, la secuenciación de ambos genomas por NGS permitió identificar seis mutaciones en la proteasa Nla-Pro de los dos aislamientos de PVY. En la Figura 60-B se observa el alineamiento de la proteasa Nla-Pro con las respectivas variaciones; estas son respecto a f\_PVY: V21I, V35F, V104I, R119K, E184G, N194D. Probablemente, las sustituciones por isoleucina y fenilalanina, de las dos primeras mutaciones, presenten un papel significativo en la actividad catalítica de la enzima al encontrarse en uno de sus dominios  $\beta$ -laminares cercanos (Figura 60-A) y estén implicadas en el desempeño de las simulaciones moleculares.



**Figura 60. Alineamiento de secuencias de la proteasa Nla-Pro en f\_PVY y t\_PVY. A)** Estructura de la proteasa de Nla-Pro de f\_PVY (verde) con las mutaciones halladas en t\_PVY (residuos coloreados por átomos en blanco). Estas son: V21I, V35F, V104I, R119K, E184G, N194D. Entre ellas, Ile 21 y Phe-35 se encuentran en uno de los dominios  $\beta$ -laminares de la proteína y, probablemente, pueden afectar la actividad catalítica (residuos coloreados por átomos en verde) de la enzima en el proceso de hidrólisis del sustrato (cian). **B)** Alineamiento de las secuencias de ambas proteasas, sin incluir la región de corte. Las mutaciones se resaltan en rojo.

Con el propósito de corroborar el efecto de las mutaciones de la proteasa entre los dos aislamientos de PVY, se realizó un análisis del movimiento local de la cadena principal y del sustrato a lo largo de los 460 ps de la simulación. Como se observa en la Figura 61-C, f\_PVY presenta un mayor movimiento local en el dominio  $\alpha_1$ - $\beta_1$ - $\beta_2$ - $\beta_3$  que t\_PVY. Esta región contiene las mutaciones V21I, V35F que al parecer no afectan la rigidez del dominio  $\beta$ -laminar de t\_PVY. Las demás mutaciones también presentaron eventos de movilidad diferentes para ambos genomas: V104I generó distorsión en el *loop* anterior a  $\beta_8$  en t\_PVY, mientras que el residuo lisina en la posición 119 probablemente esté implicado en una mayor estabilidad entre  $\beta_8$  y  $\beta_9$ . La quinta mutación, ubicada entre la región  $\beta_{13}$  y  $\alpha_3$  generó una mayor distorsión en el *loop* respecto a f\_PVY. Similarmente, N194D causó poca rigidez en el *loop* siguiente a  $\alpha_3$  que antecede  $\beta_{14}$ . Otra región importante por su cercanía al sitio activo y al sustrato está comprendida por los residuos de las posiciones 206-222. Al corresponder al extremo N-terminal de la proteasa, es de esperarse su

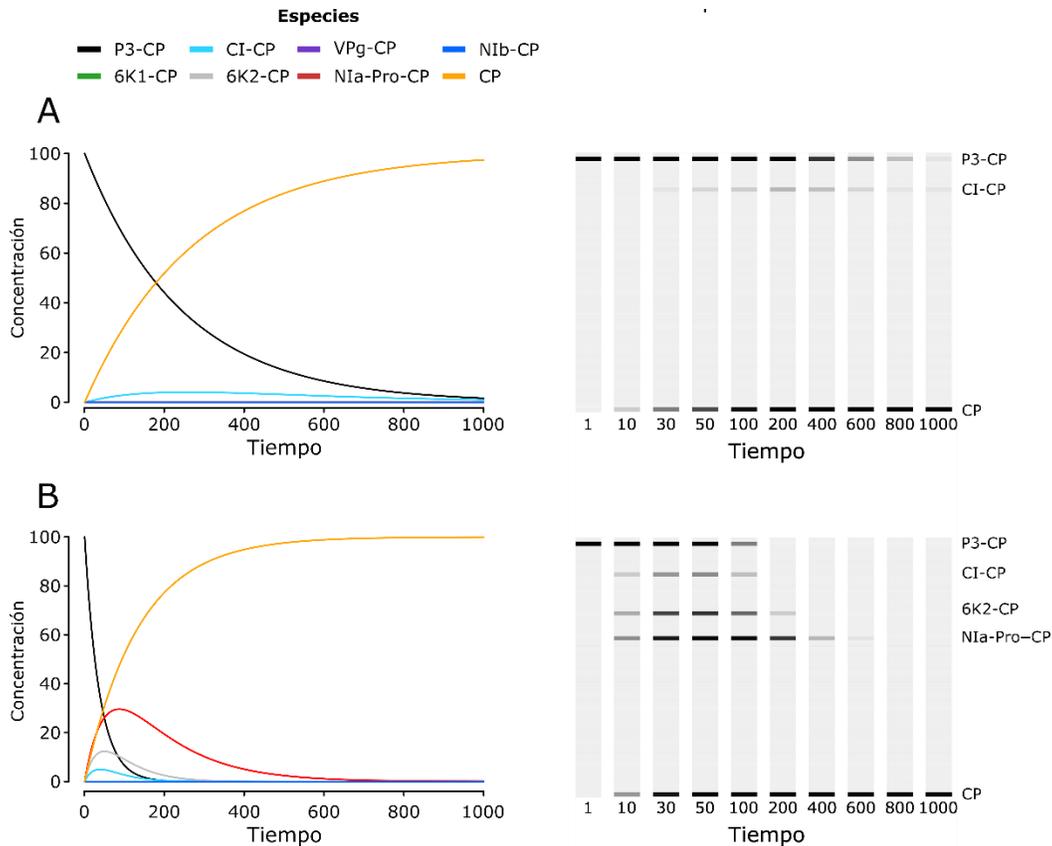
mayor movilidad local; sin embargo, en t\_PVY se presenta una mayor estabilidad pues no ocurre una distorsión observada en las respectivas posiciones de f\_PVY. Los residuos de la tríada catalítica presentan valores de RMSD similares para ambos genomas, con una ligera excepción de la histidina 46 de f\_PVY ubicada en una región de mayor movimiento en  $\alpha 2$ . Lo anterior probablemente se relacione con la eficiencia catalítica de su sitio activo, disminuyendo las probabilidades de hidrolizar el péptido que explicaría las bajas constantes cinéticas obtenidas. Adicionalmente, el análisis por movimiento local del péptido-sustrato simulado, TYEVHHQGND (Nlb/CP), también permite observar que hay una mayor rigidez en el sustrato de t\_PVY, azul oscuro y poco grosor, en comparación con f\_PVY, en azul claro y amplio grosor (Figura 61-A y B). Probablemente, esta diferencia asociada con las frecuencias de corte obtenidas en las simulaciones moleculares, tenga relación con la estabilidad general de la NIa-Pro de t\_PVY que propicia unas condiciones estructurales más favorables para la generación del ataque nucleofílico y la posterior hidrólisis peptídica. Por lo contrario, un mayor movimiento del sustrato observado en f\_PVY dificulta la orientación correcta para la ocurrencia del corte (Figura 61-D).



**Figura 61. Análisis del movimiento local de las proteasas Nla-Pro en el sistema PVY.** En **A)** y **B)** se ilustran las proteasas Nla-Pro de f\_PVY y t\_PVY, respectivamente, coloreadas según el movimiento local que presentaron durante los 460 ps de simulación. La representación empleada distorsiona el grosor de las  $\alpha$ -hélices y  $\beta$ -láminas, de acuerdo al parámetro RMSD: un grosor ancho y rojo en una sección de la molécula

significa un mayor movimiento local; mientras que entre más delgado y azul oscuro sea, habrá mayor rigidez. **C)** Alineamiento de las secuencias de la proteasa NIa-Pro con el sustrato NIb/CP de ambos PVY. Las zonas coloreadas corresponden a la escala del movimiento local observadas en A y B. Las mutaciones se muestran en color rojo y los residuos de la tríada catalítica con asteriscos. Se observa que el sustrato de t\_PVY tiene menor movimiento local en la zona del sitio activo lo que probablemente favorece su hidrólisis. **D)** Diferencia entre los valores de RMSD de ambos PVY. Un valor por encima de la media (-0.007) sugiere que la proteasa NIa-Pro de f\_PVY y su sustrato tienen un mayor movimiento local. Las regiones que no cuentan con la representación por estructura secundaria de su secuencia corresponden a los residuos móviles no detectados en la estructura cristalográfica resuelta de TEV.

Los valores obtenidos para las constantes cinéticas del sistema PVY reflejan un proceso de maduración diferente entre ambos virus. En la Figura 62-A se observa que, pasado el tiempo establecido para la simulación de la cinética, todavía no se ha logrado la desaparición del complejo P3-CP y solo el precursor CI-CP es detectado en una muy baja concentración, lo que corrobora el hecho de su baja tasa de procesamiento. Por el contrario, en t\_PVY la maduración de la poliproteína (P3-CP) se logra alrededor de 100 unidades relativas después de iniciada la cinética. Adicionalmente, se evidencian dos precursores más que en el caso anterior que alcanzan su concentración máxima en un corto tiempo al igual que CP como proteína madura (Figura 62-B). De nuevo, es importante aclarar que los sitios que presentaron frecuencias de corte iguales al 0%, dan cuenta sobre el límite de detección del modelo, y solo mediante tiempos de simulación más prolongados podría ser viable determinar dichas magnitudes.



**Figura 62. Resultados de la cinética para el sistema PVY. A)** Cinética de maduración de la poliproteína en  $f_{PVY}$ . La baja tasa de procesamiento se evidencia en la simulación de un Western blot para anti-CP. Sólo se obtuvo CI-CP como precursor y CP como proteína madura. Se observa que, transcurrido el tiempo de la cinética, todavía no se ha logrado la desaparición de la poliproteína (P3-CP). **B)** Procesamiento de la poliproteína en  $t_{PVY}$ . El proceso de maduración ocurre mucho más rápido, apreciándose tres intermediarios: CI-CP, 6K2-CP y NlIa-Pro-CP, que aparecen antes que en el caso anterior. La concentración final de la cápside también se logra al corto tiempo después de iniciada la cinética. Unidades de tiempo relativas.

La implementación de un modelo cinético de maduración que parte del genoma completo obtenido mediante datos de secuenciación masiva puede ser una herramienta útil para entender mejor la biología de los potyvirus y facilitar la interpretación de información epidemiológica y de patogenicidad. La secuenciación NGS cumple un papel clave en la obtención de genomas completos que podrán ser utilizados en la construcción de modelos por homología a partir de la secuencia aislada. Además, en conjunto con pruebas moleculares como RT-qPCR y secuenciación Sanger, es factible identificar en cuáles grupos filogenéticos se encuentran estas variantes (metodología que será explicada en los capítulos 4.2 y

4.3 del presente trabajo). Esta información, permite realizar inferencias sobre la patogenicidad de los dos aislamientos de PVY. Por ejemplo, un proceso de maduración más restringido y regulado como el de PVY<sup>NTN</sup>, evidenciado en la cinética de la Figura 62-A, podría sugerir que el virus presenta un mayor tiempo de latencia. Por lo contrario, un mecanismo de procesamiento más acelerado y de constante producción de proteínas maduras, especialmente de CP, como en la Figura 62-B, indica que el ciclo de infección del virus está dirigido a la producción masiva de viriones para aumentar el título viral con mayor velocidad. Como se mencionó anteriormente, los aislamientos de PVY corresponden a dos cepas que inducen una sintomatología diferente en la papa como su principal hospedero. Por ejemplo, para PVY<sup>N</sup> (t\_PVY) se sabe que, al ser un virus especialmente encontrado en los tejidos foliares de la planta, su mecanismo de transmisión está muy ajustado a la transmisión no persistente por áfidos, por lo que la replicación de su genoma debe presentar una tasa de procesamiento más acelerada. Adicionalmente, la rápida aparición de CP también puede estar relacionado con el mecanismo de transmisión, puesto que se han reportado interacciones entre esta proteína y HC-Pro (involucrada en la transmisión por áfidos) (Roudet-Tavert *et al.*, 2002). En cambio, la cepa recombinante PVY<sup>NTN</sup> (f\_PVY) está asociada como el agente causal de la enfermedad PTNRD de alto impacto agroeconómico. La biología de este virus sugiere que al estar destinado a los tubérculos como lugares para la replicación de su genoma, esta variante no tendría la necesidad de procesar tan rápidamente su poliproteína pues contaría con un tiempo de 4-6 meses que corresponden a los ciclos de cultivo de papa en las regiones altoandinas en Colombia. Aunque claramente, el resultado del proceso infectivo depende no sólo de las características propias del virus, como por ejemplo el procesamiento postraducciona aquí estudiado, sino de los demás componentes del tetraedro de la enfermedad (hospedante, medio ambiente, efecto antopogénico), en este trabajo se plantea que la regulación del mecanismo proteolítico mediado por NIa-Pro juega un papel clave en el título viral que alcanzan los potyvirus en los tejidos de sus hospedantes vegetales.

#### 4.1.9. Bibliografía

Anandakrishnan R, Aguilar B, Onufriev A. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research*. 2012; 40:537-541.

Berg J, Tymoczko J, Stryer L. 6 ed. Barcelona: Editorial Reverté S.A.; 2007. p. 229-249.

Cesaratto F, Burrone O, Petris G. *Tobacco etch virus* protease: a shortcut across biotechnologies. *Journal of Biotechnology*. 2016; 231:239-249.

Dougherty W, Carrington J, Cary S, Parks T. Biochemical and mutational analysis of a plant virus polyprotein cleavage site. *EMBO Journal*. 1988; 7:1281-1287.

Grzonka Z, Jankowska E, Kasprzykowski F, Kasprzykowska R, Lankiewicz L, Wiczak W, *et al.* Structural studies of cysteine proteases and their inhibitors. *Acta Biochimica Polonica*. 2001; 48:1-20.

Ivanov K, Eskelin K, Lohmus A, Mäkinen K. Molecular and cellular mechanisms underlying potyvirus infection. *Journal of General Virology*. 2014; 95:1415-1429.

Khateri H, Moarrefzadeh N, Mosahebi G, Koohi-Habibi M. Virus diseases in the tobacco fields of Guilan and Western Azerbaijan provinces of Iran. *Communications in agricultural and applied biological sciences*. 2008; 73:307-310.

Lorenzen J, Piche L, Gudmestad N, Meacham T, Shie P. A Multiplex PCR Assay to Characterize *Potato virus Y* Isolates and Identify Strain Mixtures. *Plant Disease*. 2006; 90:935-940.

Merits A, Rajamäki M-L, Lindholm P, Runeberg-Roos P, Kekarainen T, Puustinen P, *et al.* Proteolytic processing of potyviral proteins and polyprotein processing intermediates in insect and plant cells. *Journal of General Virology*. 2002; 83:1211-1221.

Murdoch D, Chow E. A Graphical Display of Large Correlation Matrices. *The American Statistician*. 1996; 50:178-180.

Onufriev A, Case D, Ullman M. A novel View of pH Titration in Biomolecules. *Biochemistry*. 2001; 40:3413-3419.

Page M, Jencks W. Entropic Contributions to Rate Accelerations in Enzymic and Intramolecular Reactions and the Chelate Effect. *Proceedings of the National Academy of Sciences of the United States of America*. 1971; 68:1678-1683.

Phan J, Zdanov A, Evdokimov A, Tropea J, Peters H, Kapust R, *et al.* Structural basis for the substrate specificity of tobacco etch virus protease. *Journal of Biological Chemistry*. 2002; 277:50564–50572.

Radisky E, Lee J, Lu C, Koshland D. Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:6835-6840.

Roudet-Tavert G, German-Retana S, Delaunay T, Delécolle B, Candresse T, Gall O. Interaction between potyvirus helper component-proteinase and capsid protein in infected plants. *The Journal of general virology*. 2002; 83:1765-1770.

Valli A, Gallo A, Calvo M, Pérez Jd.J, García, J. A novel role of the potyviral helper component proteinase contributes to enhance the yield of viral particles. *Journal of Virology*. 2014; 88:9808-9818.

Yoon H, Hwang D, Choi K, Song B. Proteolytic Processing of Oligopeptides containing the target sequences by the Recombinant *Tobacco vein mottling virus* Nia proteinase. *Molecular Cells*. 1999; 10:213-219.

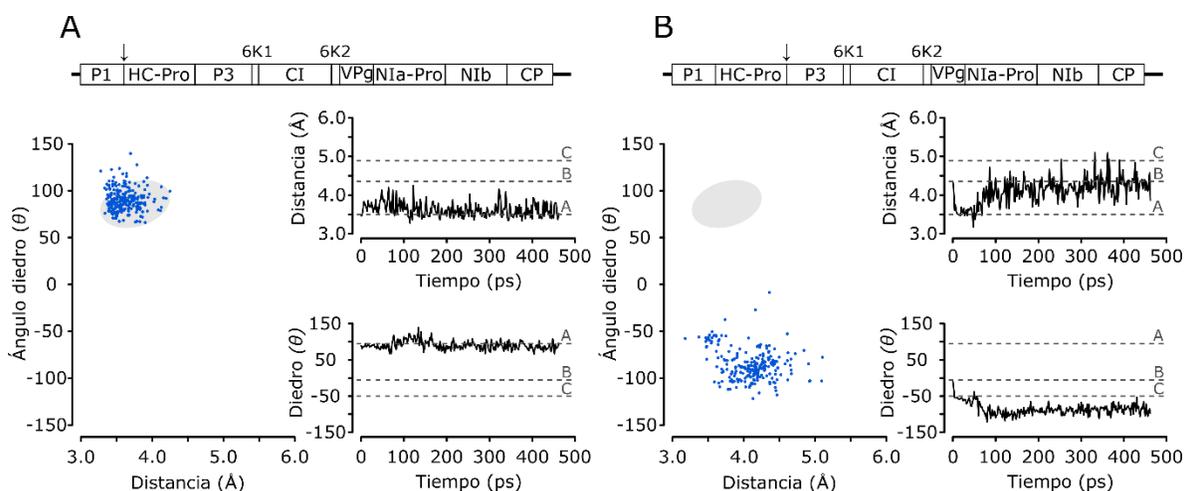
## 4.1.10. ANEXOS – Capítulo 1

### 4.1.10.1. Simulaciones moleculares en PVA

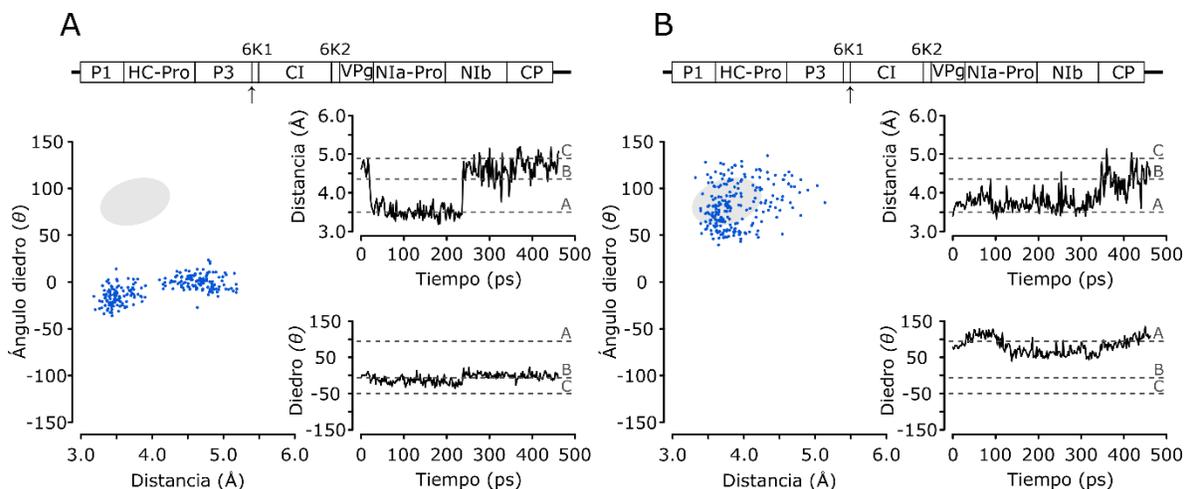
En PVA se aplicó el mismo análisis por simulaciones moleculares en las diferentes regiones de corte. Las secuencias de los péptidos-sustrato simulados, al igual que los resultados de frecuencias y mediciones, se presentan en la Tabla 6.

**Tabla 6. Resultados de las simulaciones moleculares para PVA.** En negrita se resaltan los residuos de las posiciones P<sub>-6</sub>, P<sub>-3</sub> y P<sub>-1</sub> relativas al sitio de corte (señalado por /).

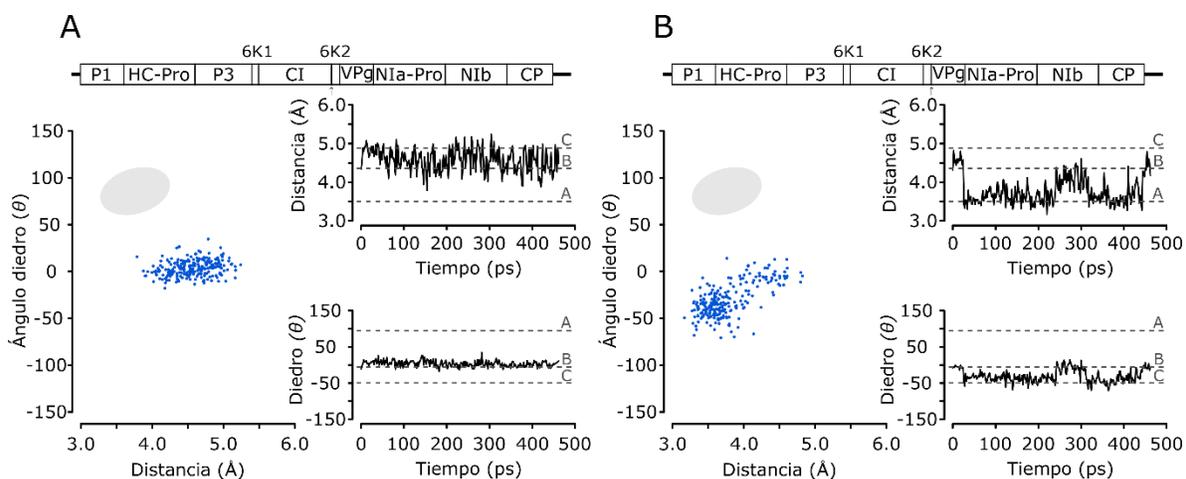
Sitio de corte	Secuencia	Frecuencia	Constante cinética	Distancia (Å)		Diedro (°)	
				$\bar{x}$	s	$\bar{x}$	s
P1/HC-Pro	MLST <b>HHY</b> /STG	0.8664	—	3.62	0.18	90.35	12.09
HC-Pro/P3	M <b>KHYR</b> VG/GTS	0.0	—	4.11	0.33	-85.56	16.32
P3/6K1	A <b>EVVLFQ</b> /AKA	0.0	k <sub>1</sub>	4.11	0.61	-7.14	11.75
6K1/CI	EN <b>TVQFQ</b> /SLD	0.4612	k <sub>2</sub>	3.89	0.36	81.56	22.47
CI/6K2	LE <b>AVQFQ</b> /GYN	0.0	k <sub>3</sub>	4.57	0.30	3.60	8.32
6K2/VPg	KG <b>VVA</b> FQ/GYN	0.0	k <sub>4</sub>	3.78	0.36	-30.45	17.98
VPg/NIa-Pro	KE <b>SVEFE</b> /STS	0.0	k <sub>5</sub>	4.46	0.38	-0.04	12.86
NIa-Pro/NIb	GE <b>AVYTQ</b> /GCD	0.4655	k <sub>6</sub>	3.79	0.22	109.06	24.35
NIb/CP	DD <b>MVYFQ</b> /AET	0.7629	k <sub>7</sub>	3.87	0.25	90.86	30.35



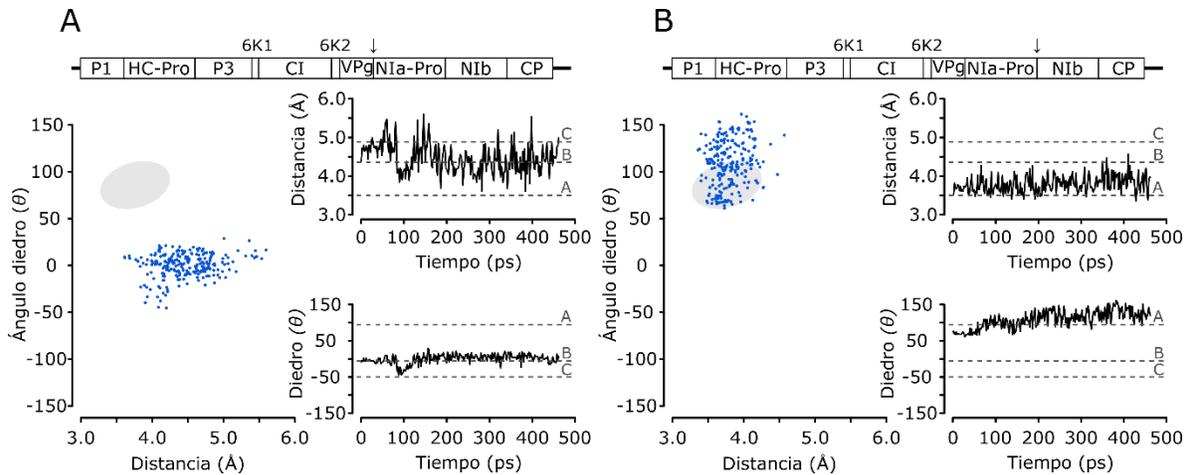
**Anexo 1. Simulación molecular de los sitios autocatalíticos en PVA. A)** Para el sitio autocatalítico P1/HC-Pro se obtuvo una frecuencia de corte del 86.64%. **B)** El otro sitio independiente de NIa-Pro presentó una frecuencia del 0%.



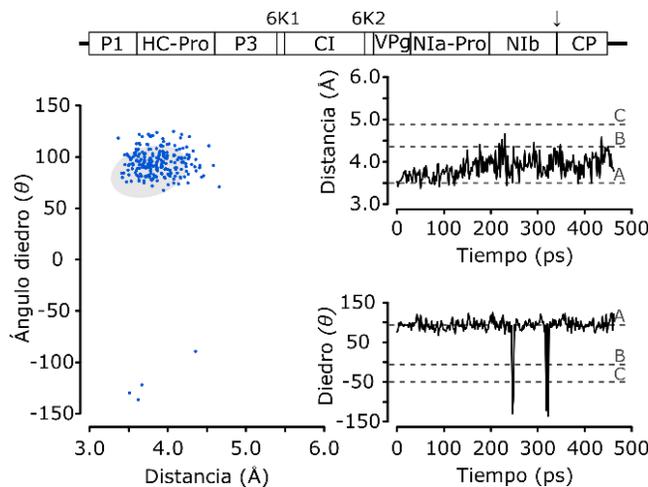
**Anexo 2. Sitios P3/6K1 y 6K1/CI de PVA.** A) Al igual que resultados obtenidos por las simulaciones moleculares en TEV y PVY, P3/6K1 tuvo una frecuencia en la región de corte del 0%. B) Para el sitio 6K1/CI se presentó una frecuencia de corte del 46.12%.



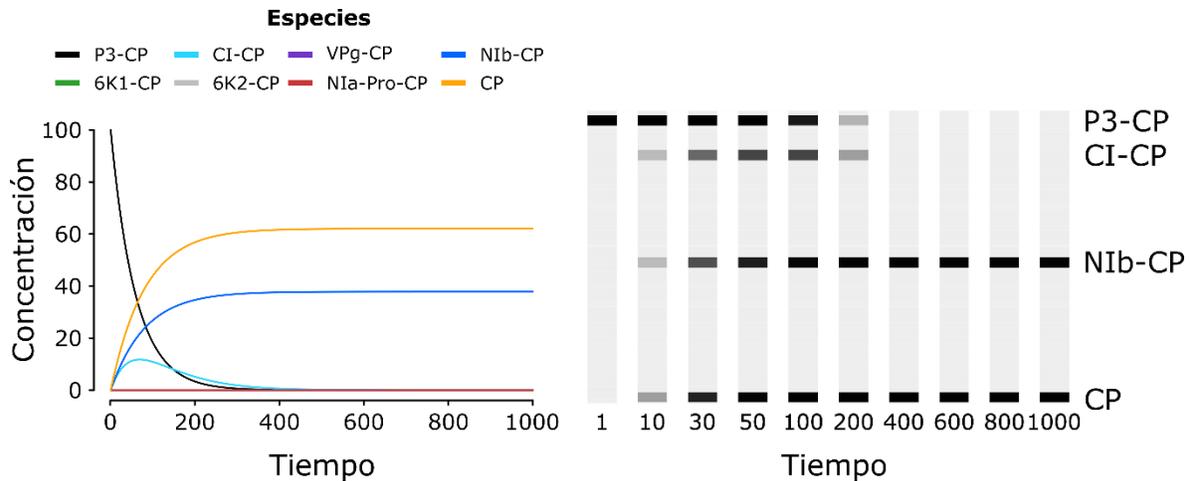
**Anexo 3. Sitios CI/6K2 y 6K2/VPg de PVA.** En ambos casos se obtuvo una frecuencia en la región de corte del 0%. El sitio A) CI/6K2 presentó una conformación más cercana a la configuración B; mientras que la región 6K2/VPg estuvo más orientada hacia la configuración C.



**Anexo 4. Sitios VPg/Nla-Pro y Nla-Pro/Nlb de PVA.** **A)** Para VPg/Nla-Pro se obtuvo una frecuencia del 0% con una orientación entre el péptido y la proteasa, más cercana a la configuración B. En **B)** el sitio Nla-Pro/Nlb tuvo una frecuencia del 46.55%, con valores de distancia  $S_V$ -C cercanos a la configuración A. Los valores del ángulo diedro presentaron un valor promedio superior al de los otros datos simulados.



**Anexo 5. Sitio Nlb/CP de PVA.** Este sitio presentó una frecuencia en la región de corte del 76.29%. Este resultado también fue observado en TEV y t\_PVY y que corresponden a los mayores valores entre los otros sitios específicos de la proteasa Nla-Pro.



**Anexo 6. Resultados de la cinética para PVA.** La maduración de la poliproteína presentó dos precursores: CI-CP y NIB-CP. Como se observa en la simulación del revelado de Western blot (anti-CP) del panel derecho, CI-CP tuvo una aparición corta (190 unidades de tiempo relativas), mientras que NIB-CP alcanzó una concentración final máxima. La poliproteína fue procesada a totalidad en la quinta parte del tiempo de simulación.

#### 4.1.10.2. Script desarrollado en R para la simulación de la cinética de especies de una poliproteína potyviral.

```
# ::: Condiciones de la simulación :::
#Tiempo (unidades relativas)
t=1000
#Definición de la matriz E(t) de 36 especies, incluyendo los
#productos finales.
C=matrix(0,36,t)
delta=0.01
#Corrección del factor entrópico "A" necesario en las
#reacciones intermoleculares. Del orden de 10-9
#Jenkins et al., 1962
A=1e-9

# ::: Matriz de constantes cinéticas :::
K=matrix(0,7,1)
K[1,1]=0
K[2,1]=0.5862
K[3,1]=0
K[4,1]=0.8664
K[5,1]=0.8664
K[6,1]=0.9095
K[7,1]=0.9397
```

```

#   ::: Concentraciones iniciales de las especies :::
C[1,1]=100    #P18
C[2,1]=0      #P17
C[3,1]=0      #P28
C[4,1]=0      #P16
C[5,1]=0      #P27
C[6,1]=0      #P38
C[7,1]=0      #P15
C[8,1]=0      #P26
C[9,1]=0      #P37
C[10,1]=0     #P48
C[11,1]=0     #P14
C[12,1]=0     #P25
C[13,1]=0     #P36
C[14,1]=0     #P47
C[15,1]=0     #P58
C[16,1]=0     #P13
C[17,1]=0     #P24
C[18,1]=0     #P35
C[19,1]=0     #P46
C[20,1]=0     #P57
C[21,1]=0     #P68
C[22,1]=0     #P12
C[23,1]=0     #P23
C[24,1]=0     #P34
C[25,1]=0     #P45
C[26,1]=0     #P56
C[27,1]=0     #P67
C[28,1]=0     #P78
C[29,1]=0     #P1
C[30,1]=0     #P2
C[31,1]=0     #P3
C[32,1]=0     #P4
C[33,1]=0     #P5
C[34,1]=0     #P6
C[35,1]=0     #P7
C[36,1]=0     #P8

#   ::: Iteración :::

#El tiempo aumenta una unidad en cada iteración del ciclo
#hasta alcanzar el tiempo establecido. De esta manera la
#concentración va aumentando (según las constantes cinéticas)
#para cada una de las 36 especies.

i=1
for (i in 1:(t-1)){

```

```

P=matrix(c(-C[1,i],-C[1,i],-C[1,i],-C[1,i],-C[1,i],
-C[1,i],-C[1,i],-C[2,i],-C[2,i],-C[2,i],-C[2,i],-C[2,i],
-C[2,i],C[1,i],C[1,i],-C[3,i],-C[3,i],-C[3,i],-C[3,i],
-C[3,i],-C[3,i],-C[4,i],-C[4,i],-C[4,i],-C[4,i],
-C[4,i],C[1,i]+C[2,i],0,C[2,i],-C[5,i],-C[5,i],-C[5,i],
-C[5,i],-C[5,i],C[3,i],0,C[1,i]+C[3,i],-C[6,i],-C[6,i],
-C[6,i],-C[6,i],-C[6,i],-C[7,i]*C[34,i]*A,-C[7,i]*C[34,i]*A,
-C[7,i]*C[34,i]*A,-C[7,i]*C[34,i]*A,C[1,i]+C[2,i]+C[4,i],0,
0,C[4,i],-C[8,i],-C[8,i],-C[8,i],-C[8,i],C[5,i]+C[3,i],
0,0,C[2,i]+C[5,i],-C[9,i],-C[9,i],-C[9,i],-C[9,i],
C[6,i],0,0,C[1,i]+C[3,i]+C[6,i],-C[10,i],-C[10,i],-C[10,i],
-C[10,i],-C[11,i]*C[34,i]*A,-C[11,i]*C[34,i]*A,
-C[11,i]*C[34,i]*A,C[7,i]*C[34,i]*A+C[4,i]+C[2,i]+C[1,i],
0,0,0,C[7,i]*C[34,i]*A,-C[12,i]*C[34,i]*A,-C[12,i]*C[34,i]*A,
-C[12,i]*C[34,i]*A,C[3,i]+C[5,i]+C[8,i],0,0,0,C[4,i]+C[8,i],
-C[13,i],-C[13,i],-C[13,i],C[6,i]+C[9,i],0,0,0,
C[2,i]+C[5,i]+C[9,i],-C[14,i],-C[14,i],-C[14,i],
C[10,i],0,0,0,C[1,i]+C[3,i]+C[6,i]+C[10,i],-C[15,i],
-C[15,i],-C[15,i],-C[16,i]*C[34,i]*A,
-C[16,i]*C[34,i]*A,
C[1,i]+C[2,i]+C[4,i]+C[7,i]*C[34,i]*A+C[11,i]*C[34,i]*A,
0,0,0,0,C[11,i]*C[34,i]*A,-C[17,i]*C[34,i]*A,
-C[17,i]*C[34,i]*A,
C[3,i]+C[5,i]+C[8,i]+C[12,i]*C[34,i]*A,
0,0,0,0,C[7,i]*C[34,i]*A+C[12,i]*C[34,i]*A,
-C[18,i]*C[34,i]*A,
-C[18,i]*C[34,i]*A,C[6,i]+C[9,i]+C[13,i],0,0,0,0,
C[4,i]+C[8,i]+C[13,i],-C[19,i],-C[19,i],C[10,i]+C[14,i],0,0,
0,0,C[2,i]+C[5,i]+C[9,i]+C[14,i],-C[20,i],-C[20,i],
C[15,i],0,0,0,0,C[1,i]+C[3,i]+C[6,i]+C[10,i]+C[15,i],
-C[21,i],-C[21,i],-C[22,i]*C[34,i]*A,
C[1,i]+C[2,i]+C[4,i]+C[7,i]*C[34,i]*A+C[11,i]*C[34,i]*A+C[16,
i]*C[34,i]*A,0,0,0,0,0,C[16,i]*C[34,i]*A,
-C[23,i]*C[34,i]*A,
C[3,i]+C[5,i]+C[8,i]+C[12,i]*C[34,i]*A+C[17,i]*C[34,i]*A,
0,0,0,0,0,C[11,i]*C[34,i]*A+C[17,i]*C[34,i]*A,
-C[24,i]*C[34,i]*A,C[6,i]+C[9,i]+C[13,i]+C[18,i]*C[34,i]*A,
0,0,0,0,0,
C[7,i]*C[34,i]*A+C[12,i]*C[34,i]*A+C[18,i]*C[34,i]*A,
-C[25,i]*C[34,i]*A,C[10,i]+C[14,i]+C[19,i],0,0,0,0,0,
C[4,i]+C[8,i]+C[13,i]+C[19,i],-C[26,i],C[15,i]+C[20,i],
0,0,0,0,0,C[2,i]+C[5,i]+C[9,i]+C[14,i]+C[20,i],
-C[27,i],C[21,i],0,0,0,0,0,
C[1,i]+C[3,i]+C[6,i]+C[10,i]+C[15,i]+C[21,i],
-C[28,i]*C[34,i]*A,
C[1,i]+C[2,i]+C[4,i]+C[7,i]*C[34,i]*A+C[11,i]*C[34,i]*A+C[16,
i]*C[34,i]*A+C[22,i]*C[34,i]*A,

```

```

0,0,0,0,0,0,C[22,i]*C[34,i]*A,
C[3,i]+C[5,i]+C[8,i]+C[12,i]*C[34,i]*A+C[17,i]*C[34,i]*A+C[23
,i]*C[34,i]*A,0,0,0,0,0,0,
C[16,i]*C[34,i]*A+C[23,i]*C[34,i]*A,
C[6,i]+C[9,i]+C[13,i]+C[18,i]*C[34,i]*A+C[24,i]*C[34,i]*A,
0,0,0,0,0,0,
C[11,i]*C[34,i]*A+C[17,i]*C[34,i]*A+C[24,i]*C[34,i]*A,
C[10,i]+C[14,i]+C[19,i]+C[25,i]*C[34,i]*A,0,0,0,0,0,0,
C[7,i]*C[34,i]*A+C[12,i]*C[34,i]*A+C[18,i]*C[34,i]*A+C[25,i]*
C[34,i]*A,C[15,i]+C[20,i]+C[26,i],0,0,0,0,0,0,
C[4,i]+C[8,i]+C[13,i]+C[19,i]+C[26,i],C[21,i]+C[27,i],0,0,0,
0,0,0,C[2,i]+C[5,i]+C[9,i]+C[14,i]+C[20,i]+C[27,i],
C[28,i]*C[34,i]*A,0,0,0,0,0,0,
C[1,i]+C[3,i]+C[6,i]+C[10,i]+C[15,i]+C[21,i]+C[28,i]*C[34,i]*
A),36,7,byrow=T)

```

```

    C[,i+1]=C[,i]+(P%*%K)*delta
}

```

```

# ::: Vector de proteínas maduras :::

```

```

#Esta instrucción es útil para establecer un rango de
#especies que posteriormente quieran ser graficadas.
#Se encuentra definido según el orden en que se definieron
#las especies.
vect.prot<-t(c(1,29,30,31,32,33,35,36))

```

```

# ::: Vector de especies que contienen CP y CP :::
vect.cp<-t(c(1,3,6,10,15,21,28,36))

```

```

# ::: Gráficas de la cinética de maduración :::

```

```

#Grafica cada uno de las especies intermediarias que
contienen CP

```

```

plot(C[1,],col="black",type="l",xlab='Tiempo',ylab='Especies
CP')
lines(C[3,],col="green",type="l")
lines(C[6,],col="cyan",type="l")
lines(C[10,],col="grey",type="l")
lines(C[15,],col="purple",type="l")
lines(C[21,],col="red",type="l")
lines(C[28,],col="blue",type="l")
lines(C[36,],col="orange")

```

```

# ::: Simulación de revelados Western blot :::

```

```

# Se inicia con la definición de una matriz de la forma:
# -10 filas que representan 10 instantes de tiempo
# -8 columnas que representan las especies (en este caso 8)
#      P18 P28 P48 P58 P68 P78 P8
#T1   [
#T2   |
#T3   |
#T4   |
#T5   |
#T6   |
#T7   |
#T8   |
#T9   |
#10   [

#Creamos la matriz:

HM<-matrix(0,10,8)

#En cada ciclo se asigna la concentración de cada una de las
#especies definidas en el vector en los diferentes 10 estados
#de tiempo. En este caso, de intermediarios de CP.
for (i in 1:8){
  HM[1,i]=C[vect.cp[1,i],1]
  HM[2,i]=C[vect.cp[1,i],10]
  HM[3,i]=C[vect.cp[1,i],30]
  HM[4,i]=C[vect.cp[1,i],50]
  HM[5,i]=C[vect.cp[1,i],100]
  HM[6,i]=C[vect.cp[1,i],200]
  HM[7,i]=C[vect.cp[1,i],400]
  HM[8,i]=C[vect.cp[1,i],600]
  HM[9,i]=C[vect.cp[1,i],800]
  HM[10,i]=C[vect.cp[1,i],1000]
}

# ::: Revelado :::

#Requiere un archivo de ponderación con cada uno de los
#pesos moleculares de las especies que se deseen graficar.

gel<-matrix(0,1000,10)
gel[195,]=HM[,1]
gel[219,]=HM[,2]
gel[233,]=HM[,3]
gel[281,]=HM[,4]
gel[288,]=HM[,5]

```

```
gel[312,]=HM[,6]  
gel[351,]=HM[,7]  
gel[509,]=HM[,8]
```

#Se toma un segmento de la matriz anterior. El efecto de una membrana para la simulación Western blot se logra con la función heat map.

```
seg=gel[190:510,]  
heatmap(seg,Colv=NA,Rowv=NA,scale='col',col=paste("gray",99:1  
,sep=""))
```

**4.2. Capítulo 2. Detección y caracterización molecular del *Potato virus Y* (PVY) en cultivos de papa del Norte de Antioquia (Colombia)**

**Artículo Publicado en:**

REVISTA DE PROTECCIÓN VEGETAL. Vol. 31 fasc. 1, p. 9-19; 2016 (ISSN: 2224-4697)

**DANIEL MUÑOZ ESCUDERO**

Laboratorio de Microbiología Industrial, Facultad de Ciencias. Universidad Nacional de Colombia - Sede Medellín, Colombia. E-mail: [dmunoze@unal.edu.co](mailto:dmunoze@unal.edu.co)

**PABLO ANDRÉS GUTIÉRREZ SÁNCHEZ, PhD.**

Laboratorio de Microbiología Industrial, Facultad de Ciencias. Universidad Nacional de Colombia - Sede Medellín, Colombia. E-mail: [paguties@unal.edu.co](mailto:paguties@unal.edu.co)

**MAURICIO MARÍN MONTOYA, PhD.**

Laboratorio de Biología Celular y Molecular, Facultad de Ciencias. Universidad Nacional de Colombia - Sede Medellín, Colombia. E-mail: [mamarinm@unal.edu.co](mailto:mamarinm@unal.edu.co)

**Correspondencia:**

Maurico Marín Montoya, PhD., Departamento de Biociencias, Facultad de Ciencias. Universidad Nacional de Colombia – Sede Medellín. Calle 59A N° 63-20, Medellín – Colombia. E-mail: [mamarinm@unal.edu.co](mailto:mamarinm@unal.edu.co)

## RESUMEN

El *Potato virus Y* (PVY) es uno de los virus más limitantes del cultivo de papa en Colombia y otros países. Es transmitido por áfidos y por tubérculo-semilla. Su manejo se fundamenta en el control de vectores y en la siembra de material certificado, para lo que se requiere de métodos de detección altamente sensibles y específicos. En este trabajo se evaluaron dos técnicas de diagnóstico de PVY (TAS-ELISA e Inmunocaptura-RT-PCR en tiempo real (IC-RT-qPCR), en tejido foliar de papa var. Diacol Capiro de tres lotes del Norte de Antioquia (Colombia). Adicionalmente, se secuenció el genoma del virus mediante secuenciación de nueva generación. El PVY fue detectado con ambas técnicas en los tres lotes, aunque éstas difirieron significativamente en su sensibilidad; con IC-RT-qPCR se encontró el virus en 88.8% de las muestras, mientras que con TAS-ELISA en 33.3% de éstas. La naturaleza de los amplicones fue confirmada por secuenciación Sanger (identidad=98-100%) y comparación con los valores de temperaturas de fusión ( $T_m$ ) del control positivo ( $77.5\text{ °C} \pm 1\text{ °C}$ ). Los análisis bioinformáticos, permitieron caracterizar el genoma completo de la cepa PVY\_Yarumal, con una profundidad de 716x y un tamaño de 9701 nt. Presenta un marco de lectura abierto (ORF) que codifica para una poliproteína de 3061 aminoácidos, flanqueada por regiones no traducidas (UTR) 5' y 3' de 187 y 328 nt, respectivamente. Los análisis filogenéticos tanto de la poliproteína como de la cápside indican su afinidad con genotipos de la raza necrosante PVY<sup>NTN</sup>, siendo además identificados los motivos K<sub>400</sub>E<sub>419</sub> y QELA asociados a dicho linaje.

**Palabras clave:** ELISA, NGS, Potyvirus, RT-qPCR, *Solanum tuberosum*.

## ABSTRACT

*Potato virus Y* (PVY) is one of the most limiting viruses in potato crops in Colombia and the rest of the world. As PVY is transmitted by aphids and tuber-seed, vector control and the use of certified seed are currently the best control strategies for this virus; however, the latter requires the availability of highly sensitive and specific detection methods. In this work, TAS-ELISA and IC-RT-qPCR were evaluated as diagnostic tools for PVY in potato using leaf samples of var. Diacol Capiro from three different potato plots in northern Antioquia (Colombia). Both techniques differed significantly in sensitivity, as 88.8% of samples tested positive for PVY using IC-RT-qPCR in contrast to only 33.3% with TAS-ELISA. Sanger sequencing of the RT-qPCR amplicons confirmed the presence of PVY and melting temperature analysis was in excellent agreement with the positive control ( $77.5\text{ }^{\circ}\text{C} \pm 1\text{ }^{\circ}\text{C}$ ). The complete genome sequence of the PVY strain infecting potato crops in northern Antioquia, PVY-Yarumal, was also obtained using Next-generation sequencing (NGS). Bioinformatic analysis resulted in a contig of 9701 nt with average sequence depth of 716x. The PVY\_Yarumal contains a ORF coding for a 3061 residue protein flanked by 5' and 3' untranslated regions of 187 and 328 nt, respectively. Phylogenetic analysis using the complete polyprotein CDS and partial CP segments suggests phylogenetic affinity with necrotic strain PVY<sup>NTN</sup>; sequence motifs K<sub>400</sub>E<sub>419</sub> and QELA typical of PVY<sup>NTN</sup> were also found in PVY-Yarumal.

**Key words:** ELISA, NGS, Potyvirus, RT-qPCR, *Solanum tuberosum*.

## INTRODUCCIÓN

El *Potato virus Y* (PVY), la especie tipo del género *Potyvirus* (Familia *Potyviridae*), es uno de los virus más limitantes para la producción de papa en el mundo, al ocasionar pérdidas hasta del 80% en los rendimientos de los cultivos y afectar la calidad de los tubérculos, cuando se presentan infecciones por cepas necrosantes como PVY<sup>NTN</sup> y PVY<sup>Wi</sup> (1, 2). Su infección causa un amplio rango de síntomas que incluyen mosaicos suaves a severos, rugosidad de hojas, defoliación, enanismo, necrosis foliar y de tubérculos y finalmente reducción en el rendimiento de las plantas (3). Este virus es transmitido de manera no persistente por al menos 100 especies de áfidos, siendo *Myzus persicae* la más eficiente; así como también por medios mecánicos y por tubérculos-semilla infectados (4). Es por esto que el establecimiento de los cultivos con material de siembra certificado por su sanidad viral es un aspecto fundamental para el manejo de este y otros virus que infectan la papa (2, 5).

Los viriones de PVY se caracterizan por presentar partículas filamentosas de 730 nm de longitud y 11 nm de diámetro, con una molécula de ARN de cadena sencilla de aproximadamente 9700 nucleótidos (nt), con una cola de poli-A en el extremo 3' y una proteína VPg unida covalentemente al extremo 5' (4). Su genoma se expresa a partir de la traducción de una poliproteína que es procesada por tres proteasas de origen viral (Nla, HC-Pro y P1) dando origen a 10 proteínas funcionales denominadas: P1, HC-Pro, P3, 6K1, CI, 6K2, Nla-VPg, Nla-Pro, NIb y CP; las dos últimas correspondientes a la replicasa viral RdRp (*RNA-dependent RNA polymerase*) de 59.7 kDa y a la proteína de la cápside (CP) de 30kDa (6).

El PVY presenta altas tasas de mutación y recombinación genética, por lo que esta especie es muy variable a nivel biológico y molecular, siendo tradicionalmente reconocidas cinco razas principales denominadas como: PVY<sup>O</sup> (raza ordinaria), PVY<sup>N</sup> (raza necrosante), PVY<sup>C</sup> (raza que causa estriados puntiformes), PVY<sup>Z</sup> (raza no necrosante) y PVY<sup>NP</sup> (no patogénica a papa), con base en sus diferencias en patogenicidad sobre papa, tabaco y pimentón (7). En la última década utilizando técnicas moleculares, se han identificado diferentes linajes genéticos asociados a

cada una de esta razas, siendo los más característicos aquellos denominados como PVY<sup>NTN</sup> y PVY<sup>N:O / Wi</sup> (3, 7) que corresponden a cepas que causan necrosis en tubérculos de papa y en particular la enfermedad denominada como "Anillo Necrótico de los Tubérculos" (*Tuber Necrotic Ringspot Disease of Potato* - PTNRD, por sus siglas en inglés) (7, 8).

Para la detección específica de PVY se utilizan diferentes metodologías biológicas, serológicas y moleculares. Entre las primeras se destacan la inspección visual de síntomas foliares asociados a la infección por este virus y las pruebas de inoculación mecánica utilizando brotes de tubérculos y plantas indicadoras, especialmente del género *Nicotiana*; proceso que requiere entre 6 y 8 semanas y obedece a criterios subjetivos dependientes de la experticia del evaluador (9). De otra parte, las metodologías serológicas como ELISA, han sido ampliamente utilizadas para la detección del PVY al ofrecerse comercialmente anticuerpos tanto monoclonales como policlonales en formatos que permiten el diagnóstico masivo de muestras; sin embargo estas metodologías tienen restricciones de sensibilidad y diferenciación de variantes, siendo especialmente limitantes para la detección de este virus en tejidos con bajo título, como por ejemplo los tubérculos en latencia (5, 10). En contraste, los métodos de detección viral basados en ácidos nucleicos como RT-PCR convencional (11, 12) y RT-PCR en tiempo real (RT-qPCR) (9, 13, 14) son altamente sensibles, de rápida ejecución y permiten el diagnóstico simultáneo de diferentes virus, cuando se realizan en formatos multiplex (6).

En los últimos años se han incorporado a las herramientas de diagnóstico y caracterización molecular de virus, las de secuenciación de nueva generación (NGS), también denominadas como de secuenciación masiva, de alto rendimiento, o de secuenciación en paralelo. Estas técnicas han modificado los esquemas de detección de patógenos no cultivables y en particular de virus y viroides, ya que no dependen del uso de cebadores, sondas o anticuerpos, al basarse en el empleo de adaptadores con secuencias conocidas y en los sistemas WS ("wash and scan"), pudiendo generar en un solo corrido desde 1 Gb hasta 1000 Gb (1 Tb) de

información de secuencias metagenómicas (15). Los sistemas de NGS más utilizados en la actualidad incluyen las plataformas 454 GS FLX+ (Roche), SOLiD (ABI) y HiSeq 2000/2500 (Illumina), que presentan diferentes fundamentos bioquímicos y de captura de datos. Los sistemas HiSeq 2000/2500, generan secuencias pareadas (*mate-paired-end*) que facilitan los análisis bioinformáticos de ensamblaje *de novo* y pueden entregar hasta 1 Tb de información por cada corrido de seis días. Estas plataformas se fundamentan en un proceso de amplificación por puentes ("*bridge PCR*") en celdas de flujo que produce millones de grupos de amplicones (*clusters*) y en el empleo posterior de dNTPs marcados con fluorocromos que actúan como terminadores reversibles en las reacciones de secuenciación, generando lecturas (*reads*) de entre 100 y 300 pb (16).

Utilizando sistemas NGS, en los últimos años se han descubierto al menos 49 virus de plantas con genomas tanto de ARN como de ADN; 36 de los cuales fueron clasificados en 16 familias virales y nueve se constituyeron en las especies tipo de igual número de nuevos géneros (16). En Colombia, el uso de metodologías NGS también ha permitido detectar y caracterizar molecularmente diferentes especies de virus de plantas solanáceas, como el *Andean potato latent virus* (APLV) (17), *Potato virus S* (PVS) (18), *Tamarillo leaf malformation virus* (TaLMV) (19), *Potato virus X* (PVX) (20) y PVY (21).

Con el fin de ofrecer alternativas de detección que apoyen los programas de manejo integrado de enfermedades virales y de producción de tubérculos-certificados de papa en Colombia y otros países, en este trabajo se evaluó la utilidad de la técnica de Inmuncaptura-RT-PCR en tiempo real (IC-RT-qPCR) para el diagnóstico de PVY en tejido foliar de plantas de papa var. Diacol-Capiro obtenidas en tres lotes del norte del departamento de Antioquia (Colombia), comparándose su eficiencia de detección con respecto a la metodología tradicionalmente empleada de TAS-ELISA. Adicionalmente, se secuenció completamente el genoma de este virus utilizando el sistema NGS Illumina HiSeq 2000, como base para su caracterización molecular en Colombia.

## MATERIALES Y MÉTODOS

### Muestras

En dos lotes de cultivo de *S. tuberosum* var. Diacol-Capiro del municipio de Santa Rosa de Osos (06°38'51"N, 75°27'37"W, altitud 2550 msnm) y en un lote ubicado en Yarumal (6°57'0"N, 75°25'1"W, altitud 2100 msnm) en el departamento de Antioquia (Colombia), se obtuvieron de manera aleatoria 15 muestras de folíolos de plantas de papa en estado de floración, con el fin de evaluar la presencia de PVY mediante pruebas de TAS-ELISA e IC-RT-qPCR. Adicionalmente, en el lote de Yarumal, se obtuvo una muestra con síntomas típicos de la infección por PVY (mosaico severo, enanismo y deformación de brotes) con el fin de ser utilizada para la secuenciación del transcriptoma foliar por NGS.

### Pruebas de TAS-ELISA e IC-RT-qPCR

Las 45 muestras foliares de folíolos de papa se evaluaron para detectar la presencia de PVY utilizando pruebas de TAS-ELISA (*Triple Antibody Sandwich*) de la compañía Agdia (Indiana, EE.UU.), que emplean anticuerpos policlonales para la captura del PVY y monoclonales conjugados a fosfatasa alcalina para su revelado, siguiendo las instrucciones del fabricante. Los resultados colorimétricos fueron cuantificados en un equipo Multiskan RC/MS/EX (Thermo, EEUU), incluyendo en cada prueba un control positivo y un control negativo, consistentes de tejidos vegetales suministrados en forma liofilizada por el proveedor comercial (Agdia). La definición de las muestras positivas para la infección de PVY, se basó en la fórmula del valor de corte (*Cut-off*) reportada por Bioreba ([www.bioreba.com](http://www.bioreba.com)), [ $Cut-off = (\text{promedio} + 3 \text{ desviaciones estándar}) \times 1.1$ ]. La naturaleza viral del control positivo de PVY fue confirmada por RT-PCR convencional y secuenciación por el método de Sanger, con los cebadores PVYCPF (5' ACC ATC AAG SAA ATG ACA CA 3') y PVYCPR (5' CGG AGA GAC ACT ACA TCA CA 3') que amplifican el gen completo (801 pb) de CP (22), siguiendo el protocolo reportado por Henao-Díaz *et al.* (23).

Una vez finalizadas las pruebas de TAS-ELISA, se procedió a la liberación en todas las muestras (tanto las positivas como negativas) de las partículas virales que putativamente fueron reconocidas por los anticuerpos específicos para PVY, utilizando el procedimiento descrito por Wetzel et al. (24), en el que se utilizan 70  $\mu$ L de buffer de liberación (Tris-HCl 10mM pH 8.0, 1% Triton X 100) e incubación a 70°C durante 10 min. Estos productos de inmunocaptura, fueron utilizados como molde para evaluar la presencia del PVY mediante RT-qPCR con SYBR Green I. Para la síntesis del ADN copia (ADNc) se utilizaron 200 U de la enzima Maxima Reverse Transcriptase (Thermo), 1X de buffer RT, 0,5 mM de dNTPs, 100 pmol del cebador Oligo-dT (se une a la cola de poli-A del extremo 3' del genoma potyviral), 20 U de inhibidor de ARNasas y 12,5  $\mu$ L del producto de liberación post-ELISA. Las reacciones se incubaron en un termociclador T3 (Biometra, Alemania) a 65°C por 5 min, seguido de 50°C por 30 min y 85°C por 5 min. El qPCR se realizó con el kit Maxima SYBR Green/ROX qPCR Master Mix (2X) (Thermo), en 25  $\mu$ L de reacción conteniendo 12,5  $\mu$ L del kit, 10  $\mu$ L de agua estéril libre de nucleasas, de 20 a 100 ng de ADNc y 0,3  $\mu$ M de los cebadores PVY-1 FP (5'CCA ATC GTT GAG AAT GCA AAA C 3') y PVY-1 RP (5'ATA TAC GCT TCT GCA ACA TCT GAG A 3') (14) que amplifican una porción de 74 pb de la región CP de PVY. Todas las reacciones de RT-qPCR incluyeron un control negativo libre de cDNA viral y un control positivo de PVY, proveído por la compañía Agdia. La amplificación se realizó en un equipo Rotor-Gene Q-5plex Platform (Qiagen, Alemania) con adquisición de fluorescencia después de cada ciclo, siendo definidas como muestras positivas aquellas que presentaron valores de Ciclo umbral (*threshold Cycle* - Ct) menores de 35 y amplicones con temperatura de fusión ( $T_m$ ) de  $\pm 1^\circ\text{C}$  con respecto al control positivo, siguiendo los criterios de Schena et al. (25). Todas las reacciones de RT-qPCR incluyeron un control negativo libre de ADNc viral. Adicionalmente, la naturaleza viral de cinco de los amplicones que cumplieron con estas condiciones y del control positivo, fue reconfirmada por secuenciación Sanger en la compañía MacroGen (Corea del Sur), previa purificación con el kit *QIAquick Gel Extraction* (Qiagen). Las secuencias obtenidas fueron editadas con el programa Mega 6.0 (26) y

comparadas con las bases de datos moleculares mediante el programa BLASTN (<http://www.ncbi.nlm.nih.gov/BLAST>). La concordancia y diferencias significativas entre los resultados ofrecidos por las pruebas de TAS-ELISA y de IC-RT-qPCR se evaluaron mediante el índice de Cohen's Kappa y la prueba de McNemar, respectivamente, utilizando el programa Graphpad (<http://graphpad.com/quickcalcs>).

### **Pruebas de RT-qPCR en *bulks***

Con el fin de evaluar la utilidad de la técnica de RT-qPCR para la detección de PVY directamente a partir de ARN total de tejido foliar de papa, se establecieron tres *bulks* de muestras por cada lote de cultivo (S1 a S5; S6 a S10 y S11 a S15), realizándose su maceración con nitrógeno líquido para proceder a la extracción del ARN total utilizando el kit *GeneJET Plant RNA Purification* (Thermo), a partir de 100 mg de muestra, siguiendo las instrucciones del fabricante. El ARN obtenido fue eluido en 50 µL de agua tratada con DEPC y determinada su concentración y pureza por lecturas de absorbancia a 260 nm y 280 nm en un equipo Nanodrop 2000C (Thermo), para su uso en pruebas confirmatorias de la presencia de PVY mediante RT-qPCR en dos pasos. Las condiciones de amplificación y de análisis de resultados fueron similares a las descritas anteriormente, así como la confirmación de la naturaleza viral de los amplicones, que en este caso se realizó para dos de las muestras.

### **Secuenciación NGS**

En la muestra foliar sintomática colectada en el municipio de Yarumal, fue confirmada la presencia de PVY mediante TAS-ELISA y RT-qPCR, tal como se describió para las muestras aleatorias. Una vez determinada la infección por este virus, se procedió a determinar la concentración y pureza del ARN total, por lecturas de absorbancia a 260 nm y 280 nm en un equipo Nanodrop 2000C (Thermo). Posteriormente, la muestra fue tratada con el kit *TruSeq Stranded Total RNA with Ribo-Zero Plant* (Illumina, EEUU) para eliminar el ARN ribosomal y luego evaluado el número de integridad del ARN (RIN - *RNA Integrity Number*) a partir de lecturas en un equipo 2100 Bioanalyzer (Agilent Technologies, EEUU). Las

librerías de ADNc se construyeron con el kit *TruSeq RNA Sample Preparation* (Illumina) y la secuenciación masiva fue realizada en un equipo HiSeq 2000 (Illumina) de la compañía Macrogen. Una vez obtenidas las secuencias, se removieron las bases con baja calidad (Phred <30) utilizando el programa SeqTK (<https://github.com/lh3/seqtk>) y los *reads* asociados a secuencias de PVY fueron identificados por análisis de mapeo con respecto a genomas de referencia utilizando el programa Bowtie2 (27) y BLASTX (28). El genoma completo del aislamiento de PVY detectado en esta planta, fue ensamblado *de novo* utilizando rutinas en *Perl* escritas por nuestro grupo para este fin.

### **Análisis filogenéticos**

Las secuencias obtenidas por NGS para el genoma completo de PVY y por Sanger para la región CP del control positivo, fueron utilizadas como base para realizar análisis filogenéticos, a partir de su alineamiento con respecto a secuencias disponibles en GenBank para cepas de PVY de referencia de las razas PVY<sup>O</sup>, PVY<sup>N/NTN</sup>, PVY<sup>N:O/Wi</sup>, PVY<sup>C</sup> y PVY<sup>Z</sup>; así como con respecto a secuencias de CP de aislamientos de este virus obtenidos previamente en cultivos de papa y tomate de árbol de Colombia. Los alineamientos se realizaron con el programa Clustal W (29) y los análisis de agrupamiento con el algoritmo de Máxima verosimilitud y el modelo GTR utilizando el programa Mega 6.0 (26). El soporte de la topología interna de las ramas del árbol fue determinado por análisis de *bootstrap* con 500 remuestreos (30).

## **RESULTADOS Y DISCUSIÓN**

### **Pruebas de TAS-ELISA e IC-RT-qPCR**

El PVY fue detectado mediante pruebas de TAS-ELISA en los tres lotes de cultivos de papa, con una incidencia del 33.3 % para las 45 muestras evaluadas (Fig. 1). La validez de las pruebas de TAS-ELISA fue confirmada por los altos valores de absorbancia obtenidos para el control positivo comercial (promedio=2,9; SD=0,42) con relación a los valores alcanzados para los controles negativos (promedio=0,10; SD=0,032); definiéndose en el histograma de

frecuencias de absorbancia a 405 nm, un nivel de 0,41, como el valor de corte del ensayo.

Las pruebas de IC-RT-qPCR utilizando como molde los productos de la post-ELISA resultaron positivas para el 88.8% de las muestras evaluadas, con valores de Ct de 20,94 a 34,81, (promedio 28,11) y de Ct=27,55 para el control positivo (Fig. 1). El análisis de las curvas de desnaturalización, arrojó un valor de  $T_m=77,38^{\circ}\text{C}$  para el control positivo y de  $T_m=77.5 \pm 1^{\circ}\text{C}$  para las muestras (Fig. 2). Estos valores son similares a los encontrados por Medina et al. (2015) en un estudio tendiente a evaluar la presencia de PVY en tubérculos de papa de las variedades Diacol-Capiro y Criolla-Colombia en el departamento de Antioquia, pues en ese caso los controles positivos presentaron valores de  $T_m$  de  $77,5^{\circ}\text{C} \pm 0,5^{\circ}\text{C}$ . Adicionalmente, la naturaleza viral de los amplicones generados en las reacciones de IC-RT-qPCR fue confirmada por secuenciación de Sanger para cinco muestras, encontrándose niveles de identidad del 98% - 100% con respecto a secuencias de aislamientos de PVY reportados en GenBank (ej. KJ946936, KJ865800 y KJ847048). Así mismo, la secuencia del amplicon del control positivo presentó 100% de identidad con accesiones de PVY como KJ946936 y KJ865800.

Estos resultados señalan diferencias evidentes entre los niveles de sensibilidad ofrecidos por ambas técnicas para la detección de virus de plantas, al presentarse 25 disparidades (53,2% de observaciones) entre éstas, y un muy bajo índice de concordancia de Cohen's Kappa ( $k= 0,14$ ;  $SE=0,06$ ; Intervalo de confianza al 95%: 0,023 a 0.258); encontrándose además diferencias significativas en la capacidad de detección de PVY por parte de ambas pruebas (McNemart=23,04,  $P<0,0001$ ). Diversos reportes en virología vegetal han registrado tales diferencias, siendo frecuente encontrar que las metodologías basadas en RT-qPCR con respecto a pruebas de ELISA, ofrecen niveles superiores de sensibilidad en el orden de  $10^4$  a  $10^8$  veces (9, 32, 33). Específicamente para PVY, Kogovsek et al. (1) en un estudio tendiente a diseñar pruebas de RT-qPCR para diferenciar aislamientos recombinantes de PVY<sup>NTN</sup>, registran que las pruebas en tiempo real presentaban sensibilidades superiores entre  $10^5$  y  $10^7$  veces con respecto a las

pruebas de ELISA comerciales; mientras que Medina et al. (31), al evaluar la capacidad de las pruebas de RT-PCR en tiempo real para el diagnóstico de este virus en tubérculos-semilla de papa en Antioquia, encontraron que esta prueba detectó el PVY en el 80,6% de las muestras, en comparación con el 37,5% detectado con pruebas de TAS-ELISA.

### **Pruebas de RT-qPCR en *bulks***

Las pruebas de RT-qPCR a partir de ARN total de tejido foliar de papa, detectaron el PVY en ocho de los nueve *bulks* de muestras evaluadas, con valores de Ct de 19,41 a 29,17 (promedio=24,56) y de Tm de 77,1°C±0,5°C; mientras que para el control positivo y para la muestra con síntomas aparentes de infección por PVY, dichos valores fueron de Ct=24,1 (Tm=77,6) y Ct=12,98 (Tm=77,1), respectivamente (Tabla 1). Las diferencias notables entre los valores de Ct alcanzados en las pruebas de RT-qPCR a partir de ARN total en comparación con IC-RT-qPCR, indican la presencia de un mayor título viral en el procedimiento que utiliza extracción de ácidos nucleicos en comparación con la cantidad de partículas virales retenidas por los anticuerpos en los pozos de las placas de ELISA, por lo que se sugiere el empleo de esta primera metodología para confirmar la presencia de PVY en muestras de mayor relevancia, como por ejemplo aquellas procedentes de bancos madres de plantas de papa, en material *in vitro* y/o de minitubérculos utilizados para la generación de semilla Super élite y élite, en muestras de importancia cuarentenaria o en materiales avanzados de mejoramiento genético por resistencia a virus; mientras que la IC-RT-qPCR se sugiere emplear en forma rutinaria como complemento o sustitución (cuando sólo se empleen anticuerpos de captura) a las pruebas de ELISA que se utilizan tradicionalmente en los esquemas de certificación de lotes de cultivo destinados a la producción de tubérculo-semilla de papa, que para el caso de Colombia incluyen la generación de semilla básica, registrada y certificada (34), así como para apoyar estudios epidemiológicos y de control de este virus.

La efectividad del uso de la práctica de siembra de semilla certificada de papa por su sanidad viral, ha sido plenamente demostrada en países como Canadá y

Estados Unidos, en los que operan estrictos sistemas estatales y privados de certificación de semilla, acompañados por el empleo de técnicas de detección altamente sensibles como RT-PCR múltiple y RT-PCR en tiempo real (5, 13, 35), siendo estimado en EEUU, que un nivel del 10% de incidencia de PVY en material de siembra de papa, representa pérdidas económicas significativas para los agricultores (36). Por su parte, Fageria et al. (9), evaluando la incidencia de PVY en lotes de cultivos de papa de New Brunswick (Canadá), reportaron que niveles de infección en tejido foliar del 0,6% (en 2009) y 2% (en 2010), condujeron al incremento en la incidencia de este virus hasta en un 20,3% (2009) y 21,9% (2010), en los tubérculos obtenidos al momento de la cosecha. Halterman et al. (5), indican que el aumento en la detección de cepas necrosantes de PVY en los cultivos de papa de EEUU y la consiguiente disminución de variantes de la raza ordinaria, han puesto nuevamente a este virus como uno de los más limitantes para la agroindustria de producción de papa. Esto debido a que las cepas necrosantes generalmente inducen síntomas foliares atenuados y difíciles de detectar visualmente, pero causan daños considerables en la calidad de los tubérculos; siendo además reportado en los últimos años, como la principal razón para el rechazo de lotes destinados a la producción de tubérculos-semilla de papa (35).

### **Secuenciación NGS**

En la planta de papa var. Diacol-Capiro con síntomas de mosaico rugoso se confirmó la infección por PVY tanto con las pruebas de TAS-ELISA (Absorbancia a 4905 nm=2,74) como mediante RT-qPCR (Ct=12,98; Tm=77,1°C). La secuenciación NGS del transcriptoma de tejido foliar, generó una librería pareada de 6.980.917 reads, para un total de 1,396,183,400 nt (Fig. 3). Los análisis de mapeo contra genomas de referencia usando Bowtie2 y de comparación con GenBank mediante BLASTX, permitieron el ensamblaje, con una profundidad promedio de 716x, del genoma completo de PVY de 9701 nt, con PVY-Yarumal como nombre del aislamiento.

Este genoma presenta un marco de lectura abierto (ORF) que codifica para una poliproteína de 3061 a.a., flanqueado por regiones no traducidas (UTR) 5' y 3' de 187 y 328 nt (excluyendo la cola de poli-A) respectivamente. Las proteínas P1, proteinasa HC-Pro, P3, PIPO, 6K1, proteína de inclusión cilíndrica (CI), 6K2, proteína de unión al genoma (VPg), proteinasa a de inclusión nuclear (Nla-Pro), replicasa viral (Nlb) y proteína de la cápside (CP) tienen 285, 455, 365, 247, 52, 634, 52, 188, 244, 519 y 287 a.a., respectivamente y los sitios de corte proteolíticos se identificaron en los residuos 285, 740, 1105, 1157, 1791, 1843, 2031, 2275 y 2794 (Fig. 3), siendo similares a los reportados por Adams et al. (37) en su análisis del clivaje de la poliproteína potyviral. El análisis de variación interna realizado entre los *reads* generados para cada segmento del genoma de PVY-Yarumal, identificó la ocurrencia de 15 mutaciones puntuales, tres de las cuales conducen a cambios de aminoácidos en la poliproteína (K30E en P1, M1121I en 6K1 y K2758R en Nlb), lo que da cuenta de las altas tasas de mutación que se han reportado en este virus en múltiples estudios moleculares (3, 4, 8).

Los motivos y regiones que caracterizan el genoma de los potyvirus fueron identificados en el genoma de la cepa PVY\_Yarumal, e incluyeron las regiones denominadas por Turpen (38) como *potyboxes* A (TCAATACAACAT) y B (TCAAACAA) entre las posiciones 16-27 y 68-76 del extremo 5' UTR; los motivos conservados en HC-Pro FRNK (residuos de a.a. 463-466) e IGN (residuos 532-534 a.a.) asociados a la supresión de silenciamiento de genes y replicación viral, respectivamente (39); la secuencia nucleotídica GA<sub>7</sub>T en la posición (2918-2926) de P3, que da origen a la proteína P3N-PIPO (247 a.a.) resultante de un cambio en el marco de lectura identificado previamente por Chung et al. (40); los motivos GSGKSTGLP (residuos 1245-1253) y DECH (residuos 1331-1334) en CI, que indican la homología de esta proteína con la superfamilia de helicasas SF2 (41); los sitios de corte de la cisteín proteasa Nla-Pro en las uniones P3/6K1, 6K1/CI, CI/6K2, 6K2/VPg, VPg/Nla-Pro, Nla-Pro/Nlb, Nlb/CP con secuencia de reconocimiento V-X-(HE)-(QE) (37); el motivo GDD conservado en las RdRp virales (residuos 2626-2628) y la secuencia de localización nuclear II (NSLII) TPISTPDGTIVKKFRGNNSGQPSTV (residuos 2567-2591) en Nlb; y finalmente el

motivo DAG (residuos 2800-2802) en la región CP, que está asociado a la transmisión por áfidos de PVY (41).

Los análisis de variación realizados entre la secuencia de PVY-Yarumal y cepas representativas de diferentes razas de PVY reportadas en el mundo, indican niveles de identidad de nt del 99,6% (99,8% en a.a.) y 99,2% (99,4% en a.a.) con respecto a aislamientos de PVY<sup>NTN</sup> de Polonia (IUNG-4) y Alemania (Linda), respectivamente; siendo las únicas regiones que presentaron valores inferiores a 99% de identidad, la 6K2 (98,7%) (con IUNG-4) y P1 (98,5%) y 6K1 (98,07%) (con Linda) (Fig. 4). Adicionalmente, la identidad del aislamiento PVY-Yarumal como miembro de la raza N de PVY y específicamente del linaje PVY<sup>NTN</sup> fue confirmada por el hallazgo en la región HC-Pro de los dos residuos K<sub>400</sub>E<sub>419</sub> (MKIFYPDVHDAELPRILVDHET) reportados por Tribodet et al. (42) como determinantes de los síntomas de necrosis de venas inducidos por aislamientos PVY<sup>N</sup>; así como por la secuencia de P1 RLRRQELATVRT reportada por Shubert et al. (3) en aislamientos del linaje PVY<sup>NTN</sup>.

Por otra parte, los niveles promedio de identidad entre la cepa PVY-Yarumal y aislamientos representativos de PVY<sup>C</sup> fueron del 85,1% (92,1% en a.a.) y del 89,1% (93,4% en a.a.) con relación a PVY<sup>O</sup>; siendo las regiones que codifican para P1 y para las proteinasas Hc-Pro y NIa-Pro las más disímiles, con niveles de identidad de 72 a 82,5% (72% a 93,4% en a.a.) y 73 a 82,1% (81% a 92% en a.a.), con respecto a PVY<sup>C</sup> y PVY<sup>O</sup>, respectivamente (Fig. 4).

### **Análisis filogenéticos**

El árbol filogenético generado a partir de la secuencia completa que codifica para la poliproteína de PVY, presentó cuatro clados principales correspondientes a los linajes que representan las razas PVY<sup>N/NTN</sup>, PVY<sup>N:O/N-Wi</sup>, PVY<sup>O</sup> y PVY<sup>C/NP</sup>, cuyos representantes provienen de diferentes países del mundo (Fig. 5A). El clado de PVY<sup>N/NTN</sup> incluyó cepas tanto del tipo principal N como de la variante PVY<sup>NTN</sup> y se presentó a su vez subdividido en dos grupos con 100% de soporte de *bootstrap*, siendo el primero de ellos el que albergó la secuencia del aislamiento PVY-Yarumal obtenida en este trabajo, en conjunto con diferentes aislamientos de PVY

procedentes de cultivos de papa de Eurasia y Norteamérica. De forma interesante, los otros dos aislamientos de PVY completamente secuenciados en Colombia y obtenidos en plantas de tomate de cultivos de Antioquia, se ubicaron en el otro subclado de PVY<sup>N/NTN</sup>, confirmándose así la ocurrencia en este país de al menos dos linajes de cepas necrosantes infectando plantas solanáceas en Colombia, situación que había sido inferida por Henao-Díaz et al. (23) en estudios basados en secuencias de CP de aislamientos de papa y tomate de árbol.

Por otra parte, el árbol filogenético realizado para CP, presentó una topología similar al anterior, aunque las cepas de PVY<sup>O</sup> y PVY<sup>N:O</sup> se ubicaron en un solo clado, dado que las regiones de recombinación entre éstas se ubican en otras regiones del genoma diferentes a CP (4). En este análisis filogenético el aislamiento PVY-Yarumal se ubicó en un subclado que incluye cepas de este virus de Colombia y de otros países del mundo, mientras que nuevamente el aislamiento de PVY de tomate reportado en GenBank (KT290511) se agrupó sólo con aislamientos de PVY de Colombia, reconfirmándose la ocurrencia de los dos genotipos necrosantes de PVY en la zona altoandina de Antioquia (Fig. 5B) Finalmente, la secuencia del control positivo comercial (Agdia) empleado en los análisis de TAS-ELISA y RT-qPCR se ubicó en el clado PVY<sup>O</sup>, lo que coincide con su origen en EEUU y confirma su utilidad en los análisis aquí realizados.

Las topologías de los árboles filogenéticos encontradas en este trabajo coinciden con los reportes de Ogawa et al. (8), en los que se utilizaron como base de análisis 149 secuencias de poliproteínas de PVY, incluyendo 20 de Japón, siendo recurrente en sus análisis la división del linaje PVY<sup>N/NTN</sup> en dos sublinajes denominados por los autores como N-1 (N-Europa) y N-2, (N-North America); sin embargo la presencia en este trabajo y en aquellos de Gil et al. (43) y Henao-Díaz et al. (23), de un tercer linaje conformado sólo por aislamientos colombianos de PVY, confirman la ocurrencia de variantes únicas de PVY en la región Andina de Colombia y sugieren la posible patogenicidad cruzada de estos aislamientos entre diferentes hospederos de plantas solanáceas como la papa, el tomate y el tomate

de árbol, lo que requiere ser evaluado experimentalmente mediante pruebas de inoculación mecánica de virus en ambiente controlados.

Los resultados de la presente investigación, en donde se detectó el PVY en 88,8% de 45 muestras foliares de plantas de papa de cultivos del Norte de Antioquia y aquellos recientemente reportados por Medina et al. (31), en los que este virus se encontró utilizando RT-qPCR en 100% de 16 muestras de tubérculos de Diacol-Capiro y en 66,6% de Criolla-Colombia procedentes del departamento de Antioquia, ponen de manifiesto la urgencia de revisar los programas de manejo de enfermedades virales en la agroindustria de papa en Colombia y muy especialmente reforzar los programas de certificación de tubérculo-semilla, con el empleo de técnicas de detección altamente sensibles como las basadas en RT-PCR en tiempo real. De gran interés resultará además la realización de pruebas biológicas que evalúen con detalle los efectos sobre rendimientos y calidad de los tubérculos de las cepas necrosantes de PVY en las variedades locales de papa cultivadas en Colombia y otros países andinos. Lo anterior debido a que hasta el momento no es claro si la enfermedad PTNRD causadas por algunas variantes de PVY<sup>NTN</sup> se presentan en los cultivos de estos países, siendo posible que la ocurrencia de síntomas necróticos en tubérculos, sean atribuidos equivocadamente a otros patógenos como *Tobacco rattle virus* (TRV), *Potato mop-top virus* (PMTV), *Pectobacterium solanacearum*, o incluso *Phytophthora infestans*.

## **AGRADECIMIENTOS**

Esta investigación se realizó con financiación de los proyectos 19438 y 26737 de la Vicerrectoría de Investigaciones de la Universidad Nacional de Colombia y por International Foundation for Science (IFS) Grant C/4634-2.

## BIBLIOGRAFÍA

1. Kogovsek P, Gow L, Pompe-Novak M, Gruden K, Foster GD, Boonham N, Ravnikar M. Single-step RT real-time PCR for sensitive detection and discrimination of *Potato virus Y* isolates. J. Virol. Methods. 2008;149:1-11.
2. Kerlan C. Potato viruses. En: Mahy BW, van Regenmortel MH (editors). Desk encyclopedia of plant and fungal virology. Academic Press, Oxford, Reino Unido; 2008;458-471.
3. Schubert J, Fomitcheva V, Sztangret-Wisniewska J. Differentiation of *Potato virus Y* strains using improved sets of diagnostic PCR primers. J Virol Methods. 2007;140:66-74.
4. Karasev AV, Gray SM. Continuous and emerging challenges of *Potato virus Y* in potato. Annu Rev Phytopathol. 2013; 51:571-586.
5. Halterman D, Charkowski A, Verchot J. Potato, viruses, and seed certification in the USA to provide healthy propagated tubers. Pest Tech. 2012;6:1-14.
6. Crosslin JM, Hamlin LL. Standardized RT-PCR conditions for detection and identification of eleven viruses of potato and Potato spindle tuber viroid. Am J Potato Res. 2011;88:333-338.
7. Singh RP, Valkonen JPT, Gray SM, Boonham N, Jones RAC, Kerlan C, Schubert J. Brief review: The naming of *Potato virus Y* strains infecting potato. Arch Virol. 2008;153:1-13.
8. Ogawa T, Nakagawa A, Hataya T, Ohshima K. The genetic structure of populations of *Potato virus Y* in Japan; based on the Analysis of 20 full genomic sequences. J Phytopathol. 2012;160:661-673.
9. Fageria MS, Singh M, Nanayakkara U, Pelletier Y, Nie X, Wattie D. Monitoring current-season spread of *Potato virus Y* in potato fields using ELISA and real-time RT-PCR. Plant Dis. 2013; 97:641-644.

10. Crosslin JM. PVY: An old enemy and a continuing challenge. *Am J Potato Res.* 2013;90:2-6.
11. Singh RP, Nie X. Multiple virus and viroid detection and strain separation via multiplex reverse transcription-polymerase chain reaction. *Can J Plant Pathol.* 2003;25:127-134.
12. Lorenzen JH, Piche LM, Gudmestad NC, Meacham T, Shiel P. A multiplex PCR assay to characterize *Potato virus Y* isolates and identify strain mixtures. *Plant Dis.* 2006;90:935-940.
13. Agindotan BO, Shiel PJ, Berger PH. Simultaneous detection of potato viruses, PLRV, PVA, PVX and PVY from dormant potato tubers by TaqMan real-time RT-PCR. *J Virol Methods.* 2007;142:1-9.
14. Singh M, Singh RP, Fageria MS, Nie X, Coffin R, Hawkins G. Optimization of a Real-Time RT-PCR assay and its comparison with ELISA, conventional RT-PCR and the grow-out test for large scale diagnosis of *Potato virus Y* in dormant potato tubers. *Am J Potato Res.* 2013;90:43-50.
15. Roossinck MJ, Martin DP, Roumagnac P. Plant virus metagenomics: Advances in virus discovery. *Phytopathology.* 2015;105:716-727.
16. Wu Q, Ding S, Zhang Y, Zhu S. Identification of viruses and viroids by Next-Generation Sequencing and homology dependent and homology independent algorithms. *Annu Rev Phytopathol.* 2015;53:1-20.
17. Kreuze J, Koenig R, De Souza J, Vetten HJ, Muller G, Flores B, Ziebell H, Cuellar W. The complete genome sequences of a Peruvian and a Colombian isolate of *Andean potato latent virus* and partial sequences of further isolates suggest the existence of two distinct potato-infecting tymovirus species. *Virus Res.* 2013;173:431-435.
18. Gutiérrez PA, Alzate JF, Marín M. Complete genome sequence of a novel *Potato virus S* strain infecting *Solanum phureja* in Colombia. *Arch Virol.* 2013;158:2205-2208.

19. Gutiérrez PA, Alzate JF, Marín M. Genome sequence of a virus isolate from tamarillo (*Solanum betaceum*) in Colombia: evidence for a new potyvirus. Arch Virol. 2014;160:557-560.
20. Gutiérrez PA, Alzate JF, Marín M. Complete genome sequence of an isolate of *Potato virus X* (PVX) infecting Cape gooseberry (*Physalis peruviana*) in Colombia. Virus Genes. 2015;50:518-522.
21. Villamil-Garzón A, Cuellar WJ, Guzmán-Barney M. Natural co-infection of *Solanum tuberosum* crops by the *Potato yellow vein virus* and potyvirus in Colombia. Agron Colomb. 2014;32:213-223.
22. Glais L, Tribodet M, Kerlan C. Genomic variability in Potato potyvirus Y (PVY): evidence that PVYNW and PVYNTN variants are single to multiple recombinants between PVYO and PVYN isolates. Arch Virol. 2002;147:363-378.
23. Henao-Díaz E, Gutiérrez-Sánchez P, Marín-Montoya M. Análisis filogenético de aislamientos del *Potato virus Y* (PVY) obtenidos en cultivos de papa (*Solanum tuberosum*) y tomate de árbol (*Solanum betaceum*) en Colombia. Actual Biol. 2013;35:219-232.
24. Wetzel T, Candresse T, Macquaire G, Ravelonandro M, Dunez J. A highly sensitive immunocapture polymerase chain reaction method for plum pox potyvirus detection. J Virol Methods. 1992;39:27-37.
25. Schena L, Nigro F, Ippolito A, Gallitelli D. Real-time quantitative PCR: a new technology to detect and study phytopathogenic and antagonistic fungi. Eur J Plant Pathol. 2004;110:893-908.
26. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis. Mol Biol Evol. 2013;30:2725-2729.
27. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357-359.
28. Gish W, States DJ. Identification of protein coding regions by database similarity search. Nat Genetics. 1993;3:266-272.

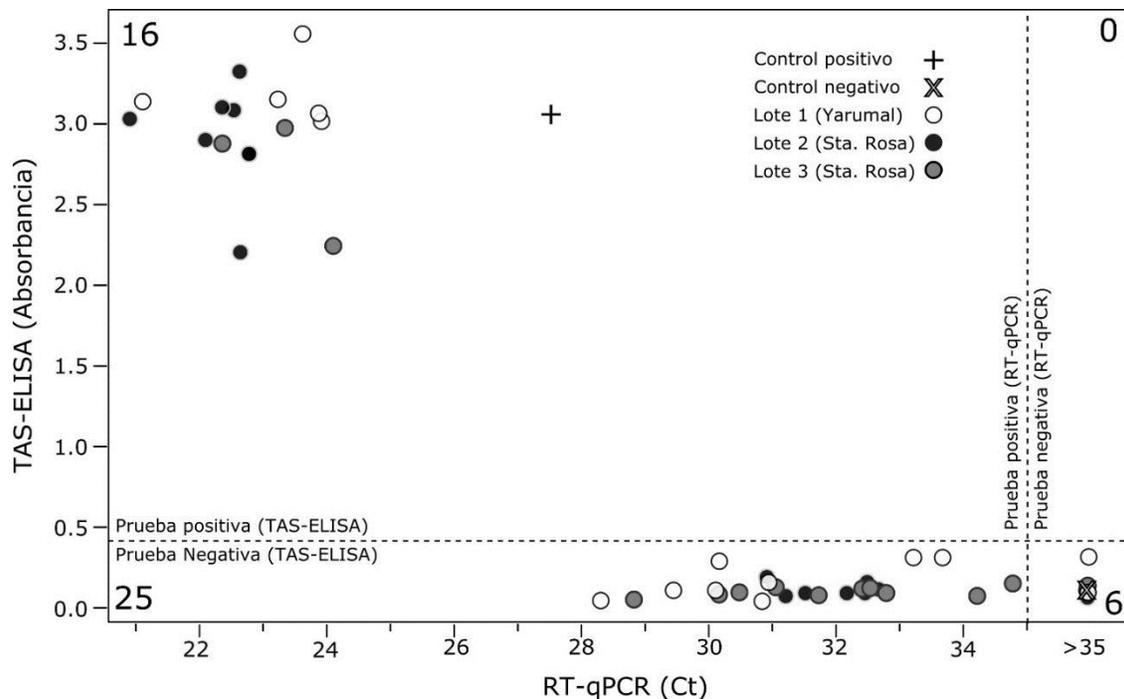
29. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23:2947-2948.
30. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985;39: 783-791.
31. Medina HC, Gutiérrez PA, Marín M. Detección del *Potato virus Y* (PVY) en tubérculos de papa mediante TAS-ELISA y qRT-PCR en Antioquia (Colombia). *Bioagro*. 2015;27:83-92.
32. Mumford R, Walsh K, Barker I, Boonham N. Detection of *Potato mop top virus* and *Tobacco rattle virus* using a multiplex realtime fluorescent reverse-transcription polymerase chain reaction assay. *Phytopathology*. 2000;90:448-453.
33. Bertolini E, Moreno A, Capote N, Olmos A, de Luis A, Vidal E, Pérez-Panadés J, Cambra M. Quantitative detection of *Citrus tristeza virus* in plant tissues and single aphids by real-time RT-PCR. *Eur J Plant Pathol*. 2008;120:177-188.
34. Instituto Colombiano Agropecuario-ICA. 2003. Resolución 2501. Requisitos mínimos para la producción de semilla certificada de papa. Disponible: <http://www.ica.gov.co/getattachment/0c8d645e-90cd-4537-9cad-1828e0a0653a/2003R2501.aspx>. (Acceso Julio de 2015).
35. Frost K, Groves RL, Charkowski AO. Integrated control of potato pathogens through seed potato certification and provision of clean seed potatoes. *Plant Dis*. 2013;97:1268-1280.
36. Nolte P, Whitworth J, Thornton MK, McIntosh CS. Effect of seedborne *Potato virus Y* on performance of Russet Burbank, Russet Norkotah, and Shepody potato. *Plant Dis*. 2004;88:248-252.
37. Adams MJ, Antoniw JF, Beaudoin F. Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. *Mol Plant Pathol*. 2005;6:471-487.
38. Turpen T. Molecular cloning of a *Potato virus Y* genome: nucleotide sequence homology in non-coding regions of potyviruses. *J Gen Virol*. 1989;70:1951-1960.

39. Shibolet Y, Haronsky E, Leibman D, Arazi T, Wassenegger M, Whitham SA, Gaba V, Gal-On A. The conserved FRNK box in HC-Pro, a plant viral suppressor of gene silencing, is required for small RNA binding and mediates symptom development. *J Virol.* 2007;81:13135-13148.
40. Chung BY, Miller WA, Atkins JF, Firth AE. An overlapping essential gene in the *Potyviridae*. *Proc Natl Acad Sci USA.* 2008;105: 5897-5902.
41. Revers F, García JA. Molecular biology of potyviruses. *Adv Virus Res.* 2015;92:101-199.
42. Tribodet M, Glais L, Kerlan C, Jacquot E. Characterization of *Potato virus Y* (PVY) molecular determinants involved in the vein necrosis symptom induced by PVYN isolates in infected *Nicotiana tabacum* cv. Xanthi. *J Gen Virol.* 2005;86:2101–2105.
43. Gil JF, Cotes JM, Marín M. Incidencia de potyvirus y caracterización molecular de PVY en regiones productoras de papa (*Solanum tuberosum* L.) de Colombia. *Rev Col Biotecnol.* 2011;85-93.

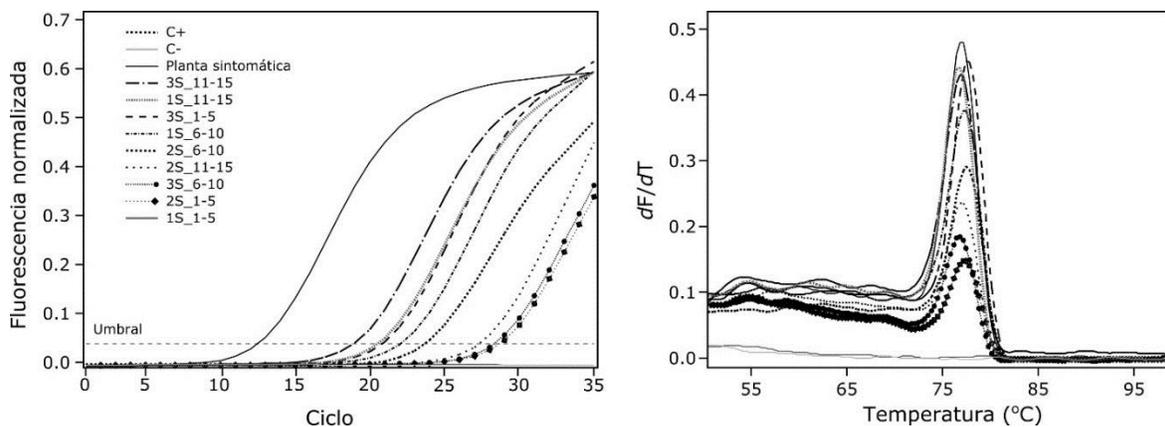
**TABLA 1.** Confirmación por RT-qPCR de la presencia del *Potato virus Y* (PVY) en tejidos foliares de plantas de papa procedentes del norte del departamento de Antioquia (Colombia) / Confirmation by RT-qPCR of the presence of *Potato virus Y* (PVY) in potato leaf tissue from samples collected from northern Antioquia (Colombia).

Muestra	Procedencia	Valor de Ct*	Valor de Tm**
<b>IC-RT-qPCR (Post-ELISA)</b>			
<b>C-</b>		>35	
<b>C+ (Agdia)</b>	EEUU	24,1	77,6
<b>Sintomática</b>	Yarumal	12,98	77,1
<b>1S_1-5</b>	Yarumal	>35	
<b>1S_6-10</b>	Yarumal	22,56	77,2
<b>1S_11-15</b>	Yarumal	21,3	77,5
<b>2S_1-5</b>	Santa Rosa	29,17	77,4
<b>2S_6-10</b>	Santa Rosa	26,43	77,1
<b>2S_11-15</b>	Santa Rosa	27,79	76,9
<b>3S_1-5</b>	Santa Rosa	21,01	76,7
<b>3S_6-10</b>	Santa Rosa	28,81	76,7
<b>3S_11-15</b>	Santa Rosa	19,41	76,9

\*Ct corresponde al ciclo umbral de amplificación; Ct>35 representa resultados negativos en las pruebas de RT-qPCR. \*\*Tm indica la temperatura de fusión, es decir aquella en la que el 50% de la doble hélice de los amplicones se encuentran desnaturalizados.

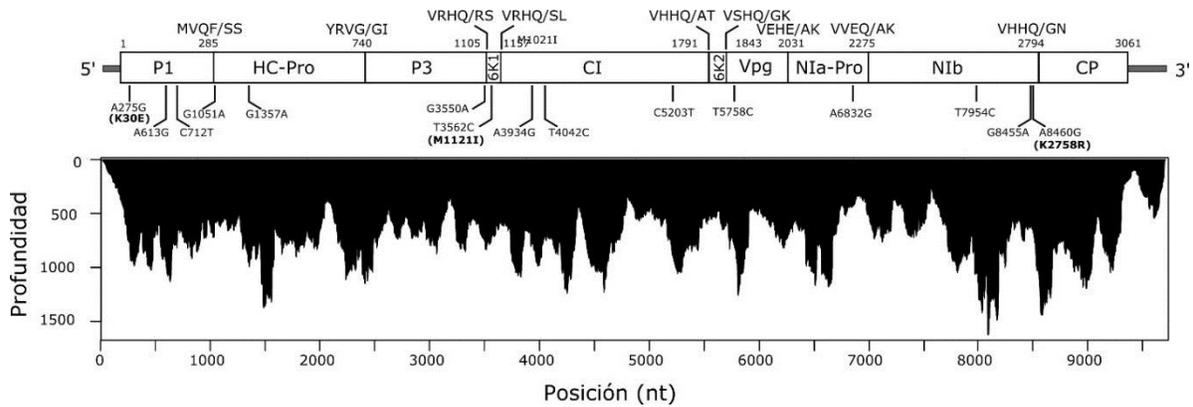


**Figura 1.** Resultados de las pruebas de TAS-ELISA e IC-RT-qPCR para la detección del *Potato virus Y* (PVY) en tejidos foliares de papa, obtenidos en el norte del Departamento de Antioquia (Colombia). En cada cuadrante se presenta el total de muestras que resultaron positivas o negativas para cada prueba. / Correlation plot between TAS-ELISA and IC-RT-qPCR detection of *Potato virus Y* (PVY) in leaf samples from northern Antioquia (Colombia). Quadrants are delimited based on the threshold values considered positive for each method.

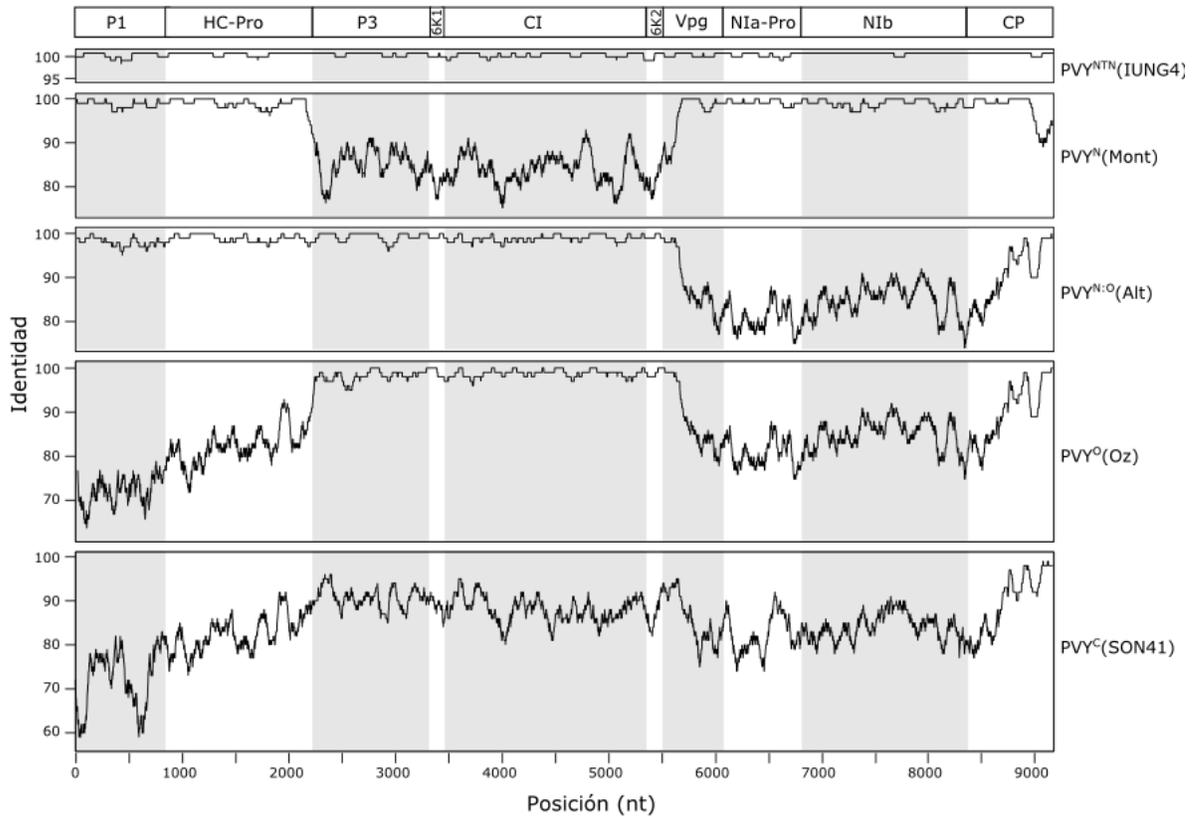


**Figura 2.** (Izq) Curvas de amplificación por RT-qPCR utilizando el sistema SYBR Green I y los cebadores PVY-1 FP y PVY-1 RP para la detección del *Potato virus Y* (PVY) en tejidos foliares de papa obtenidos en el norte del Departamento de Antioquia (Colombia). (Der) Perfiles de la curva de desnaturalización de amplicones específicos de PVY obtenidos por RT-qPCR. / (Left) RT-qPCR

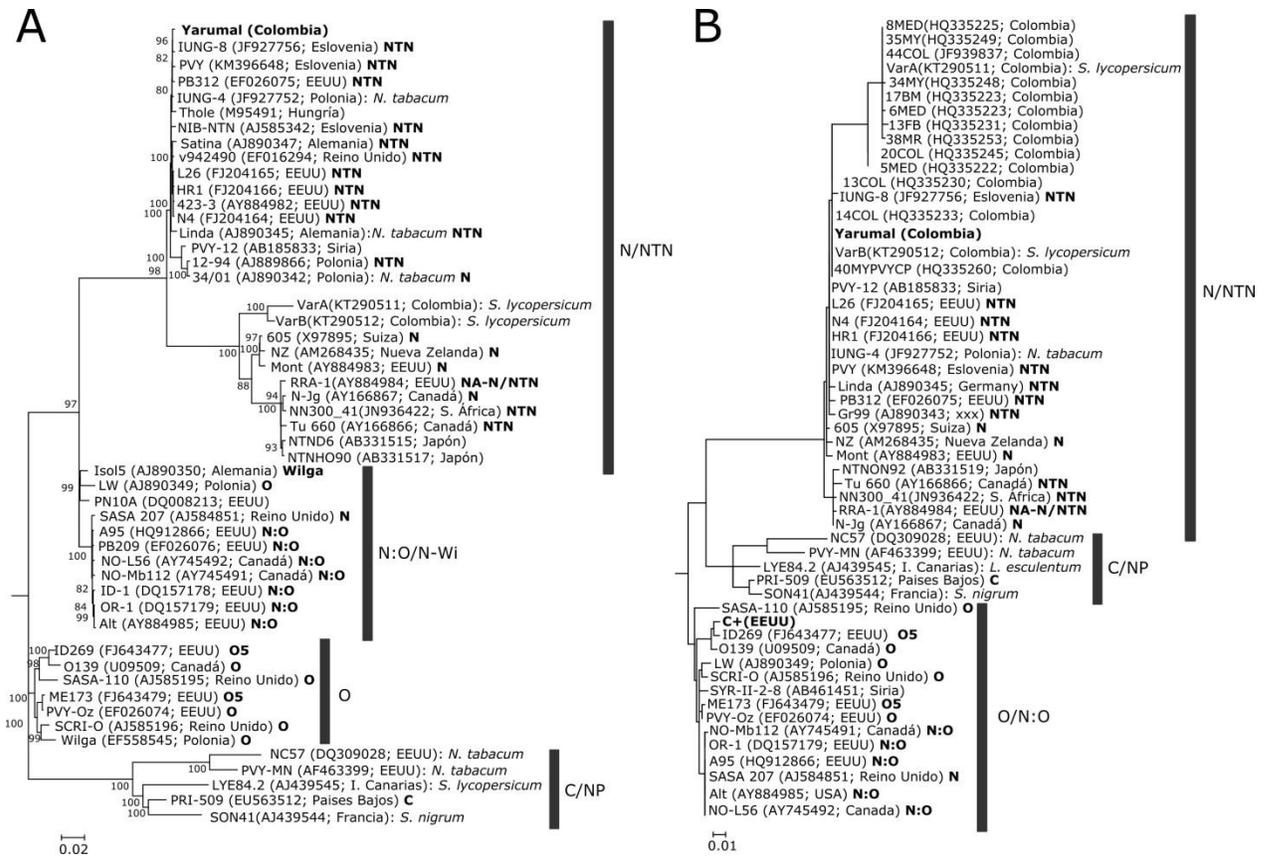
Amplification profiles using SYBR Green I and primers PVY-1 FP and PVY-1 RP of potato leaf samples from northern Antioquia (Colombia). (Right) Melting curves of the RT-qPCR amplicons shown on the left.



**Figura 3.** Profundidad de secuenciación obtenida mediante NGS para el genoma de *Potato virus Y* cepa PVY\_Yarumal procedente de una planta sintomática de papa var. Diacol-Capiro del Norte de Antioquia (Colombia). En la parte superior de la barra representando el genoma de PVY, se presentan las posiciones de aminoácidos y las señales de corte de la poliproteína por proteasas virales; mientras que en la parte inferior se indican los sitios polimórficos, señalándose en negrita aquellos que conducen a cambios de aminoácidos. / Sequence depth of *Potato virus Y* strain PVY\_Yarumal as determined by NGS sequencing of a symptomatic potato var. Diacol-Capiro plant from northern Antioquia (Colombia). The upper panel shows the protease cleavage sites of the mature protein (top) as well as polymorphic sites (below). Mutations resulting in amino acid changes are shown in bold.



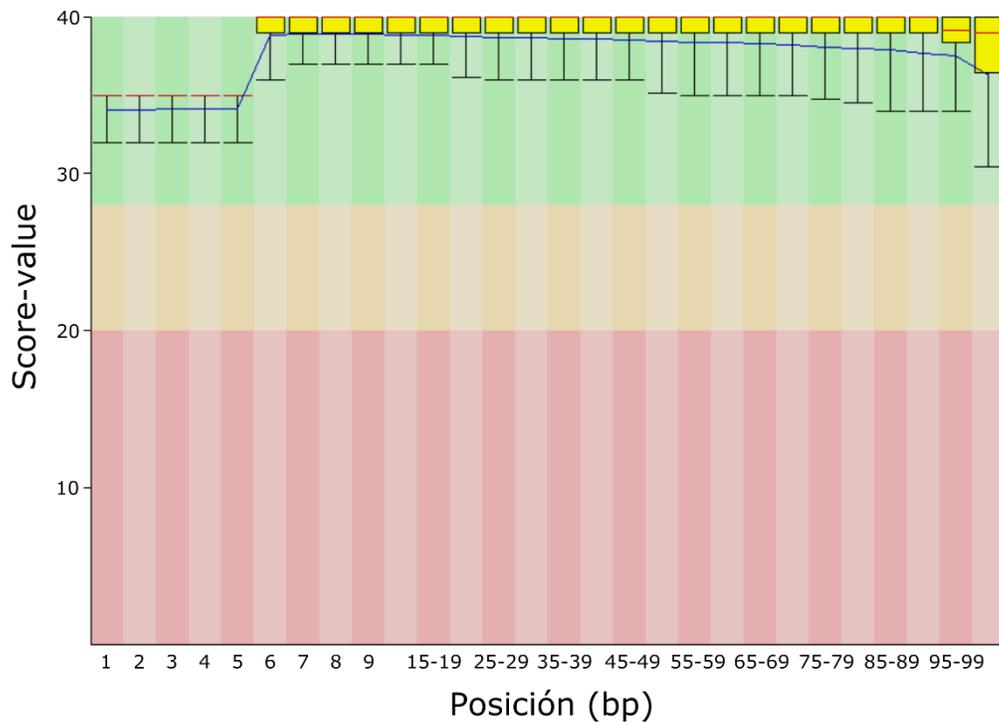
**Figura 4.** Niveles de identidad entre la cepa de *Potato virus Y* y cepa PVY\_Yarumal y otros aislamientos de PVY que representan diferentes razas de este virus en el mundo. / Local percent identity of the *Potato virus Y* strain PVY-Yarumal CDS with respect to other PVY isolates representing all different races.



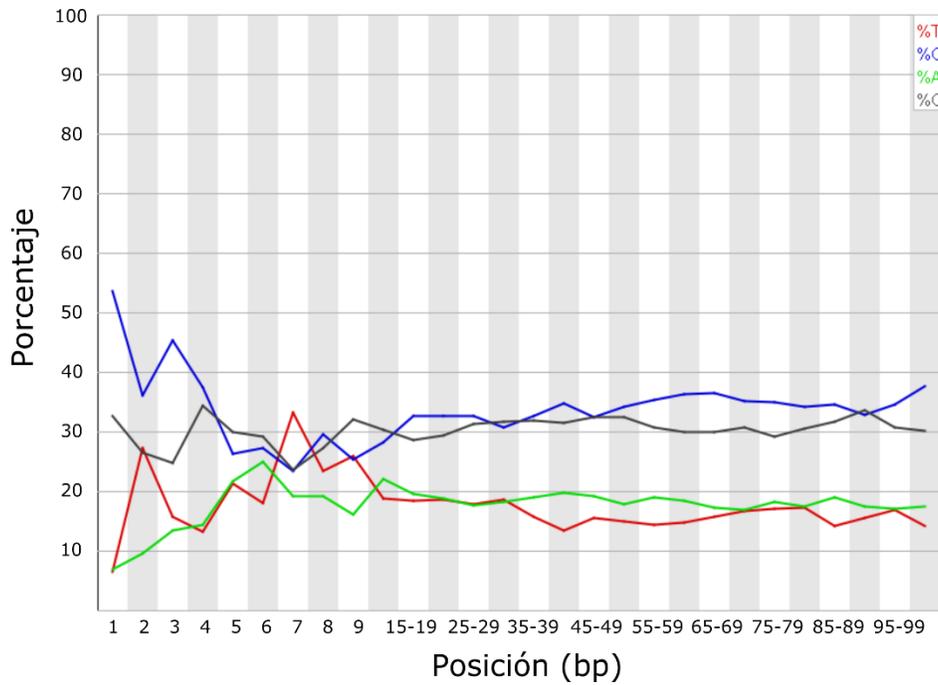
**Figura 5. (A)** Árbol filogenético basado en secuencias de la regiones que codifican para la poliproteína y **(B)** la cápside viral de aislamientos del *Potato virus Y* (PVY) procedentes de cultivos de papa del norte de Antioquia (Colombia) y otros lugares del mundo. Los números sobre las ramas indican los valores de *bootstrap* >80% y las denominaciones externas se refieren a las razas que representan los clados principales definidos en los análisis filogenéticos. / Maximum-likelihood phylogenetic trees using the complete polyprotein coding region **(A)** and a segment of the coat protein **(B)** of *Potato virus Y* strains from northern Antioquia (Colombia) and other countries. Bootstrap values >80% are shown at branch points and clades labelled according to their race.

## 4.2.1. ANEXOS – Capítulo 2

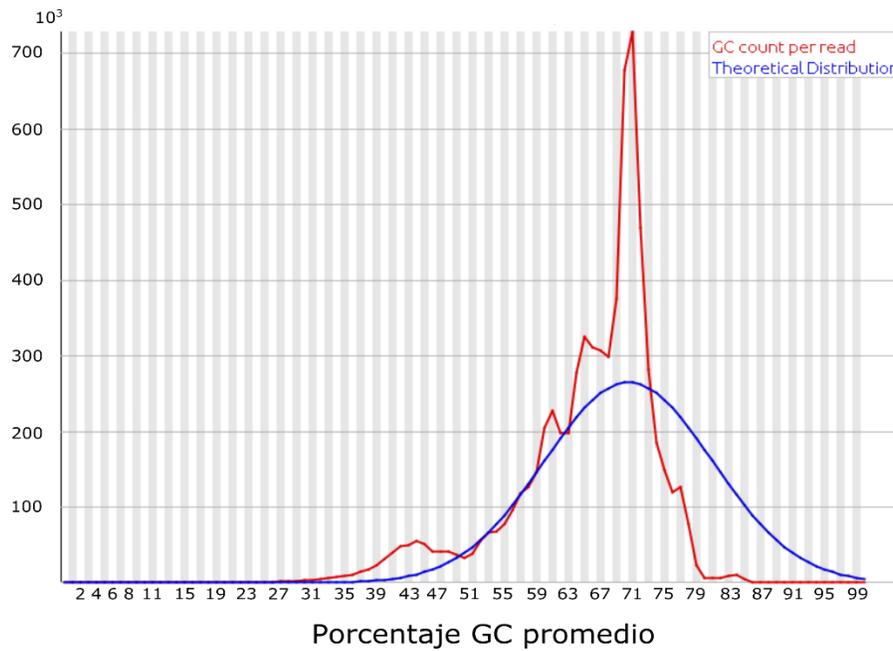
### 4.2.1.1. Análisis de los datos provenientes de la secuenciación HiSeq 2000 (Illumina)



**Anexo 1.** Los valores de calidad de los *reads* se representan mediante un diagrama Box-Whisker. La línea azul indica la calidad promedio. El fondo de la gráfica divide el eje de las ordenadas en tres zonas: buena calidad (verde), calidad razonable (naranja) y mala calidad (roja). En la mayoría de sistemas de secuenciación, es común observar un decaimiento hacia el final de las secuencias.

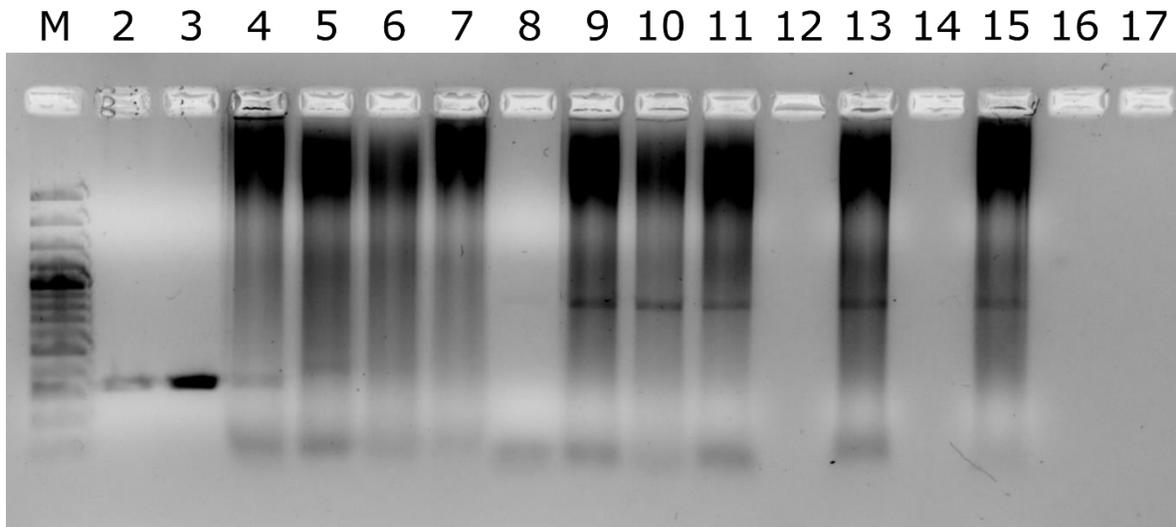


**Anexo 2.** Proporción de los nucleótidos a lo largo de la longitud de los *reads*.

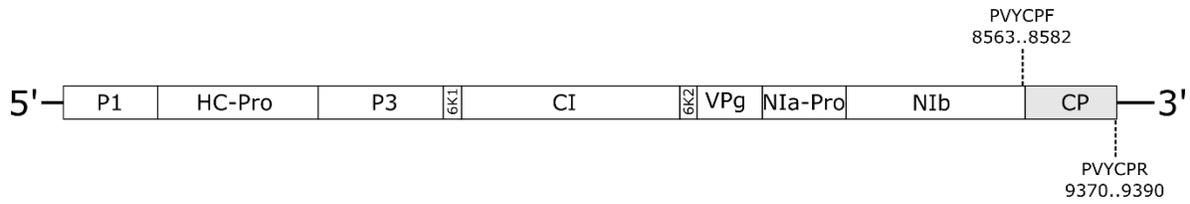


**Anexo 3.** Distribución del contenido de guanina-citosina (GC) a lo largo de toda la secuencia (rojo) en comparación con una distribución normal (azul). El pico central corresponde al contenido GC del genoma.

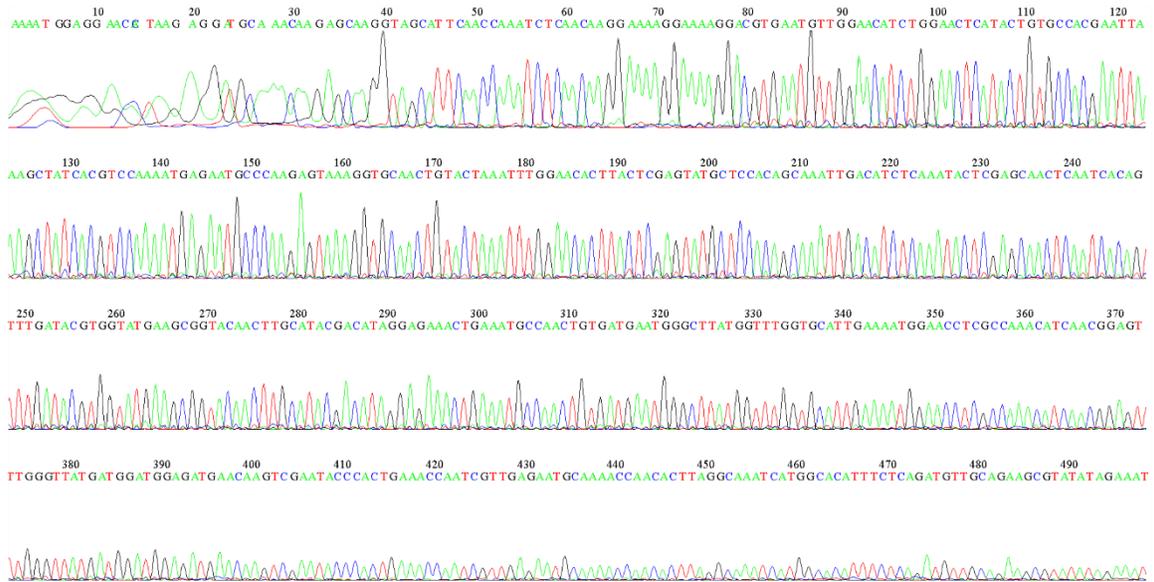
**4.2.1.2. Gel de purificación, secuenciación Sanger, electroferogramas y resultados BLAST**



**Anexo 4.** Las muestras de los tres lotes se distribuyen entre los carriles 8, 9, 10, 11, 13 y 15 (*primers* PVYCPF – PVYCPR, con un tamaño de banda de 827 pb), para su posterior secuenciación por Sanger. El marcador de peso molecular de 100pb (Thermo Scientific) se ubica en el carril M.



File: 147mar\_PVYCPF.ab1 Run Ended: 2015/4/4 3:7:26 Signal G:1582 A:1721 C:1417 T:1316  
 Sample: 147mar\_PVYCPF Lane: 57 Base spacing: 14.277169 806 bases in 9651 scans Page 1 of 2 



**Anexo 5.** Electroferograma de la región codificante para la cápside (CP) de PVY secuenciada mediante los *primers* PVYCPF y PVYCPR (panel superior).

Description	Max score	Total score	Query cover	E value	Ident	Accession
Potato virus Y isolate Yarumal_varB, complete genome	1283	1283	100%	0.0	100%	<a href="#">KT336552.1</a>
Potato virus Y isolate Yarumal, complete genome	1283	1283	100%	0.0	100%	<a href="#">KT336551.1</a>
Potato virus Y isolate La_Union_4 capsid protein gene, partial cds	1283	1283	100%	0.0	100%	<a href="#">KR902701.1</a>
Potato virus Y isolate La_Union_2 capsid protein gene, partial cds	1283	1283	100%	0.0	100%	<a href="#">KR902699.1</a>
Potato virus Y isolate La Union polyprotein gene, complete cds	1283	1283	100%	0.0	100%	<a href="#">KR149260.1</a>
Potato virus Y isolate 444 coat protein gene, partial cds	1283	1283	100%	0.0	100%	<a href="#">KJ741115.1</a>
Potato virus Y isolate 410 coat protein gene, partial cds	1283	1283	100%	0.0	100%	<a href="#">KJ741107.1</a>
Potato virus Y isolate 40MPVYCP coat protein (CP) gene, partial cds	1283	1283	100%	0.0	100%	<a href="#">HQ335260.1</a>
Potato virus Y isolate FI:50 7a:06 polyprotein gene, partial cds	1278	1278	100%	0.0	99%	<a href="#">JX432987.1</a>
Potato virus Y isolate FI:413 2b:06 polyprotein gene, partial cds	1278	1278	100%	0.0	99%	<a href="#">JX432982.1</a>
Potato virus Y isolate FI:V213 16:07 polyprotein gene, partial cds	1278	1278	100%	0.0	99%	<a href="#">JX432978.1</a>
Potato virus Y isolate FI:V3 17:07 polyprotein gene, partial cds	1278	1278	100%	0.0	99%	<a href="#">JX432976.1</a>
Potato virus Y isolate FI:V12 2:07 polyprotein gene, partial cds	1278	1278	100%	0.0	99%	<a href="#">JX432974.1</a>
Potato virus Y isolate FI:V322 12:07 polyprotein gene, partial cds	1278	1278	100%	0.0	99%	<a href="#">JX432971.1</a>
Potato virus Y isolate 3D, complete genome	1278	1278	100%	0.0	99%	<a href="#">KJ946936.2</a>
Potato virus Y strain NTN, complete genome	1278	1278	100%	0.0	99%	<a href="#">KM396648.1</a>

**Anexo 6.** Resultados del BLASTN usando como *query* la secuencia obtenida por Sanger de la región de la cápside (primers PVYCPF – PVYCPR).

Description	Max score	Total score	Query cover	E value	Ident	Accession
Potato virus Y isolate VarA, complete genome	60.8	60.8	95%	5e 07	95%	<a href="#">KT290511.1</a>
Potato virus Y isolate mar7, complete genome	60.8	60.8	95%	5e 07	95%	<a href="#">KR270797.1</a>
Potato virus Y strain 44COL capsid protein gene, partial cds	60.8	60.8	95%	5e 07	95%	<a href="#">JF939837.1</a>
Potato virus Y isolate 41MYPVY nonfunctional coat protein (CP) gene, partial sequence	60.8	60.8	95%	5e 07	95%	<a href="#">HQ335262.1</a>
Potato virus Y isolate 42444648MYPVY nonfunctional coat protein (CP) gene, partial sequence	60.8	60.8	95%	5e 07	95%	<a href="#">HQ335259.1</a>
Potato virus Y isolate 38MR PVY coat protein (CP) gene, partial cds	60.8	60.8	95%	5e 07	95%	<a href="#">HQ335253.1</a>
Potato virus Y isolate 35MY PVY coat protein (CP) gene, partial cds	60.8	60.8	95%	5e 07	95%	<a href="#">HQ335249.1</a>
Potato virus Y isolate 34MY PVY coat protein (CP) gene, partial cds	60.8	60.8	95%	5e 07	95%	<a href="#">HQ335248.1</a>
Potato virus Y isolate 20COL PVY coat protein (CP) gene, partial cds	60.8	60.8	95%	5e 07	95%	<a href="#">HQ335245.1</a>
Potato virus Y isolate 17BM PVY coat protein (CP) gene, partial cds	60.8	60.8	95%	5e 07	95%	<a href="#">HQ335238.1</a>
Potato virus Y isolate 13FB PVY coat protein (CP) gene, partial cds	60.8	60.8	95%	5e 07	95%	<a href="#">HQ335231.1</a>
Potato virus Y isolate 7MED PVY coat protein (CP) gene, partial cds	60.8	60.8	95%	5e 07	95%	<a href="#">HQ335224.1</a>
Potato virus Y isolate 6MED PVY coat protein (CP) gene, partial cds	60.8	60.8	95%	5e 07	95%	<a href="#">HQ335223.1</a>
Potato virus Y isolate BR2.1 coat protein gene, partial cds	57.2	57.2	92%	7e 06	95%	<a href="#">KJ746452.1</a>
Potato virus Y isolate SL16, complete genome	55.4	55.4	95%	2e 05	92%	<a href="#">KX713170.1</a>
Potato virus Y isolate VE6 coat protein gene, partial cds	55.4	55.4	95%	2e 05	92%	<a href="#">KX253972.1</a>

**Anexo 7.** Resultados del BLASTN para la secuencia obtenida por Sanger de la región de la cápside con los primers PVY-1 FP – PVY-1 RP.

**4.3. Capítulo 3. Genome characterization of *Potato virus Y* strains infecting potato (*Solanum tuberosum* L.) in La Unión (Antioquia, Colombia)**

**Caracterización del genoma de aislamientos de *Potato virus Y* infectando papa (*Solanum tuberosum* L.) en La Unión (Antioquia, Colombia)**

**Artículo Publicado en:**

REVISTA AGRONOMÍA COLOMBIANA. Vol. 34, p. 317-328; 2016 (ISSN: 2357-3732)

**DANIEL MUÑOZ ESCUDERO**

Laboratorio de Microbiología Industrial, Facultad de Ciencias. Universidad Nacional de Colombia - Sede Medellín, Colombia. E-mail: [dmunoze@unal.edu.co](mailto:dmunoze@unal.edu.co)

**PABLO ANDRÉS GUTIÉRREZ SÁNCHEZ, PhD.**

Laboratorio de Microbiología Industrial, Facultad de Ciencias. Universidad Nacional de Colombia - Sede Medellín, Colombia. E-mail: [paguties@unal.edu.co](mailto:paguties@unal.edu.co)

**MAURICIO MARÍN MONTOYA, PhD.**

Laboratorio de Biología Celular y Molecular, Facultad de Ciencias. Universidad Nacional de Colombia - Sede Medellín, Colombia. E-mail: [mamarinm@unal.edu.co](mailto:mamarinm@unal.edu.co)

**Correspondencia:**

Mauricio Marín Montoya, PhD., Departamento de Biociencias, Facultad de Ciencias. Universidad Nacional de Colombia – Sede Medellín. Calle 59A N° 63-20, Medellín – Colombia. E-mail: [mamarinm@unal.edu.co](mailto:mamarinm@unal.edu.co)

## Abstract

*Potato virus Y* (PVY) is one of the most severe viruses affecting the production of potato (*Solanum tuberosum*) in the world. This study presents a detailed molecular analysis using next generation sequencing (NGS), IC-RT-qPCR and RT-PCR on the PVY isolates infecting seed-tubers and foliage of potato plants cv. Diacol-Capiro in La Unión (Antioquia, Colombia). Analysis of incidence by IC-RT-qPCR in 15 random leaf samples of three cultivation plots and fifteen sprouting tuber eye-buds reveal infection levels between 13.4 and 80%; a higher incidence of 86.7% was observed in seed-tuber samples with threshold cycle (Ct) values as low as 24.3. Genome assembly from a bulk of foliage samples resulted in a consensus PVY genome (PVY\_LaUnionF) of 9,702 nt and 399 polymorphic sites within the polyprotein ORF; while the assembled genome from sprouts of tubers has 9,704 nt (PVY\_LaUnionT) and contained only six polymorphic nucleotide sites. Phylogenetic analysis demonstrates that the PVY isolates from leaf samples are in the recombinant PVY<sup>NTN</sup> group (sequence identity >99%); while those from tuber sprouts are in the PVY<sup>N/NTN</sup> group with identities above 95%. Sanger sequencing of viral capsid suggests the presence of a third variant related to PVY<sup>O</sup>, a prevalent strain reported in potato fields worldwide.

**Key words:** ELISA, genomics, plant viruses, potyviruses.

## Resumen

El *Potato virus Y* (PVY) es uno de los virus más severos que afectan la producción de papa (*Solanum tuberosum*) en el mundo. Este trabajo presenta un análisis molecular detallado utilizando secuenciación de nueva generación (NGS), ICRT-qPCR y RT-PCR, de aislamientos de PVY procedentes de tubérculo-semilla y tejido foliar de plantas de papa cv. Diacol-Capiro en el municipio de La Unión (Antioquia, Colombia). Las evaluaciones de incidencia de PVY utilizando IC-RT-qPCR en 15 muestras aleatorias de tres lotes, indican niveles variables de infección de este virus (13,4 a 80%), mientras que su ocurrencia fue muy alta (86,7%) en 15 muestras de brotes de tubérculo-semilla, con valores de Ct (ciclo

umbral) tan bajos como 24,3. El ensamblaje del genoma viral obtenido a partir de una mezcla de muestras de tejido foliar, resultó en una secuencia consenso (PVY\_LaUnionF) de 9.702 nt y 399 sitios polimórficos dentro del marco de lectura abierto (ORF) que codifica para la poliproteína. El genoma ensamblado a partir de la mezcla de brotes de tubérculo-semilla presentó un tamaño de 9.704nt (PVY\_LaUnionT) y tan sólo seis sitios polimórficos. Los análisis filogenéticos indican que los aislamientos procedentes de tejido foliar hacen parte del grupo recombinante de PVY<sup>NTN</sup> (porcentaje de identidad >99%), mientras que los aislamientos de brotes de tubérculos se ubicaron en el clado PVY<sup>N/NTN</sup> con porcentajes de identidad superiores al 95%. La secuenciación de Sanger de la cápside viral detectó una tercera variante relacionada con PVY<sup>O</sup>, una raza con altos niveles de prevalencia en cultivos de papa del mundo.

**Palabras clave:** ELISA, genómica, potyvirus, virus de plantas

## Introduction

*Potato virus Y* (PVY), the type species of the genus *Potyvirus* (*Potyviridae*), is one of the most important viruses affecting the production of potato (*Solanum tuberosum* L.) worldwide (Karasev and Gray, 2013). Foliar symptoms associated to PVY can result in yield reductions between 40% to 70%, while tuber damage significantly compromise its quality and marketability (Nolte *et al.*, 2004; Fageria *et al.*, 2013). PVY can be transmitted mechanically, by aphids such as (*Myzus persicae*) in a non-persistent manner or vertically by infected tuber-seeds (Karasev and Gray, 2013). Symptoms induced by PVY are highly dependent on the type of strain and host cultivar and may include leaf mosaics, mottle and crinkling, vein necrosis and necrotic spots; some variants can also induce necrotic ringspots in the tuber skin (Quenouille *et al.*, 2013). PVY can also affect other Solanaceous crops such as tobacco (*Nicotiana tabacum*), pepper (*Capsicum annumm*) and tomato (*Solanum lycopersicum*) (Scholthof *et al.*, 2011).

PVY virions consists of flexuous particles of about 730 nm in length and a diameter of approximately 11 nm and are composed of a positive-sense single stranded RNA of approximately 9000 nt encapsidated by roughly 2000 coat proteins subunits (CP), (Ivanov *et al.*, 2014); a covalently linked virus-encoded protein (VPg) is bound to the 5' end (Revers and García, 2015). The PVY genome encodes for a polyprotein of 3060-3063 amino acids, which after processing by three viral encoded proteases, gives rise ten mature protein products: P1, HC-Pro, P3, 6K1, CI, 6K2, NIa-VPg, NIa-Pro, NIb and CP (Revers and García, 2015). Recent studies have shown that potyvirus contain two additional ORFs named PIPO (Pretty Interesting Potyviridae ORF) and PISPO (Pretty Interesting Sweet potato Potyvirus ORF) expressed as a result of RNA polymerase transcriptional slippage (Olsper *et al.*, 2015; Untiveros *et al.*, 2016).

The PVY group consists of a vast number of strains differentiated by the symptoms induced on specific hosts, serological typing and/or nucleotide sequence (Karasev and Gray, 2013). Based on the pathogenicity in potato, tobacco and pepper, PVY has been classified into five strains: PVY<sup>O</sup>, PVY<sup>N</sup>, PVY<sup>C</sup>, PVY<sup>Z</sup> and PVY<sup>NP</sup> (Singh *et al.*, 2008). PVY<sup>C</sup> and PVY<sup>O</sup> comprise strains that can induce a hypersensitive resistance response (HR) in potato cultivars with the *Nc* and *Ny* genes, respectively (Karasev and Gray, 2013). PVY strains that do not induce HR genes and can cause systemic necrosis in tobacco belong to the PVY<sup>N</sup> group, while strains that fail to induce necrosis in tobacco and do not produce the symptoms typical of PVY<sup>O</sup> and PVY<sup>C</sup> in their corresponding cultivars are classified as PVY<sup>Z</sup> (Singh *et al.*, 2008). Strains non-pathogenic to potato are referred to as PVY<sup>NP</sup> (Blanco-Urgoiti *et al.*, 1996). However, during the past decades genome sequencing has revealed that many PVY strains such as PVY<sup>N:O</sup>, PVY<sup>N-Wi</sup> and PVY<sup>NTN</sup> are the result of recombination between PVY<sup>O</sup> and PVY<sup>N</sup> (Karasev and Gray, 2013). Recombinant strains have received increased attention as they are associated with tuber necrotic ringspot disease in potato (PVY<sup>NTN</sup>) and tobacco veinal necrosis (PVY<sup>N:O</sup>, PVY<sup>N-Wi</sup>) (Ogawa *et al.*, 2012).

Due to the emerging importance of PVY, several studies have been conducted in the potato producing regions of Colombia that demonstrated the global presence of PVY<sup>N</sup> and PVY<sup>NTN</sup> in addition to two divergent PVY clades (Gil *et al.*, 2011; Gallo *et al.*, 2012; Henao *et al.*, 2013). A recent study using TAS-ELISA and RT-qPCR, revealed widespread PVY infection in *S. tuberosum* tuber-seed used as planting material in the municipalities of La Unión and Yarumal in the province of Antioquia (Medina *et al.*, 2015). Further work confirmed the generalized presence of PVY infecting potato crops in Yarumal and Next-generation sequencing (NGS) resulted in the assembly of a complete PVY genome with phylogenetic affinity to necrotic PVY<sup>NTN</sup> strains (Muñoz *et al.*, 2016). The aim of this study was to perform a detailed molecular analysis on the PVY strains infecting tuber-seed and potato cv. Diacol-Capiro plots in La Unión using ELISA, RT-qPCR, phylogenetic analysis and NGS. The presence of two different PVY genotypes with phylogenetic affinities to the PVY<sup>N</sup> and PVY<sup>NTN</sup> is demonstrated and their relationship to previously sequenced genome discussed; Sanger sequencing also suggests the presence PVY<sup>O</sup> in this region.

## **Materials and methods**

### **Plant samples and IC-RT-qPCR**

Immunocapture (IC) was conducted following the method of Wetzel *et al.* (1992) using the Triple Antibody Sandwich PVY detection kit from Agdia (Elkhart, IN). Fifteen random leaf samples collected in three (A, B and C) commercial potato cv. Diacol-Capiro fields and fifteen tuber-seed samples from storage cellars in La Unión (Antioquia) were evaluated. Positive and negative controls consisted of lyophilized plant tissue purchased from Agdia (Elkhart, IN). For cDNA synthesis, 12.5 µL of sample released after TAS-ELISA using Tris-HCl 10 mM pH 8.0 and 1 % Triton X 100 were used as template. The reaction also included 200 U of Maxima Reverse Transcriptase, 1X RT buffer, 0,5 mM dNTPs, 100 pmol Oligo-dT primers and 20 U of RNase inhibitor (Thermo Fisher Scientific, Waltham, MA) and was incubated at 65°C for 5 min, followed by 50°C for 30 min and 85°C for 5 min in a T3 thermal cycler (Biometra, Germany). For the qPCR step, the Maxima SYBR

Green/ROX qPCR Master Mix (2X) kit (Thermo Fisher Scientific, Waltham, MA) was used in 25  $\mu$ L of reaction containing 12.5  $\mu$ L of mix, 10  $\mu$ L DEPC water, specific primers PVY-1 FP (5'CCA ATC GTT GAG AAT GCA AAA C 3') and PVY-1 RP (5'ATA TAC GCT TCT GCA ACA TCT GAG A 3') (Singh *et al.*, 2013) at 0.3  $\mu$ M and 20-100 ng cDNA. Amplification cycles consisted of 10 min at 95°C to activate the polymerase, followed by 35 cycles at 95°C for 15 s and 53°C for 45 s, using a Rotor-Gene Q-5plex Platform (Qiagen, Hilden, Germany); fluorescence was measured after each amplification cycle. The samples were considered positive if they exhibited fluorescence values higher than the threshold before the 35th cycle (Schna *et al.*, 2004). Primer specificity was verified by High Resolution Melting in the 50 and 99°C range and the identity of five amplicons, including the positive control, was confirmed by Sanger sequencing.

### **Next-generation Sequencing and Sanger sequencing**

High-throughput sequencing of the *S. tuberosum* cv. Diacol-Capiro transcriptome was performed on a sample from a bulk of leaves with symptoms typically associated with PVY infection (e.g. rugose mosaic) and previously tested using TAS-ELISA with the Triple Antibody Sandwich PVY detection kit from Agdia (Elkhart, IN). Samples were ground using liquid nitrogen followed by RNA extraction with the GeneJET Plant RNA Purification Mini Kit (Thermo Fisher Scientific, Waltham, MA). The library was constructed with the TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA) and rRNA depleted with the TruSeq Stranded Total RNA with Ribo-Zero Plant kit (Illumina, San Diego, CA). Sequencing was performed with the Illumina HiSeq 2000 system service provided by Macrogen (Seoul, South Korea). A similar procedure was followed for the tuber transcriptome, in this case, a bulk sample consisting of ten tuber sprouts from different Diacol-Capiro potato seed-tubers were used. Adapter sequences and low quality bases were removed with SeqTK (<https://github.com/lh3/seqtk>). Sequence assembly was performed with Trinity (Grabherr *et al.*, 2011) and contigs corresponding to the PVY genome identified by a local BLASTN search. Both PVY genomes assemblies were confirmed by mapping with Bowtie2 (Langmead and

Salzberg, 2012) and checked for inconsistencies and assembly artifacts with Tablet (Milne *et al.*, 2010). Consensus sequences were deposited in GenBank under accession codes KR149260 (LaUnionF) and KX531041 (LaUnionT). Additionally, the CP (801 nt) region was amplified by RT-PCR in six symptomatic foliage samples (LU1, LU2, LU3, LU4, LU5 and LU6) from La Unión using primers CP PVYCPF (5'ACC ATC AAG SAA ATG ACA CA 3') and PVYCPR (5' CGG AGA GAC ACT ACA TCA CA 3') (Glais *et al.*, 2002) following the protocol reported by Henao *et al.* (2013) and sequenced by the Sanger method (GenBank accession codes KR902698-KR902703) after gel purification using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific, Waltham, MA).

### **Bioinformatic analysis**

ORFs codifying for viral proteins were identified with BLASTX (Gish and States, 1993). The phylogenetic analysis was performed in MEGA6 (Tamura *et al.*, 2013) by the Neighbor-Joining method with 1000 bootstrap replicates using the Tamura 3-parameter method (Tamura, 1992). Rate variation among sites was modeled with a gamma distribution (shape parameter = 0.3). Evolutionary analyses were conducted in MEGA6 (Tamura *et al.*, 2013). The rates of non-synonymous substitutions per non-synonymous site (Ka) and synonymous substitution per synonymous site (Ks) were estimated by the Nei-Gojobori method using a sliding window of 14 codons (Nei and Gojobori, 1986). The recombination analysis was performed by measuring the local nucleotide percent identity along the genome using in-house perl scripts and plotted with a sliding window of 100 nucleotides; sequences PVY<sup>N</sup> (AY8849983), PVY<sup>O</sup> (EF026074) and PVY<sup>C</sup> (AJ439544) were used as parents. Multidimensional scaling analysis was performed with cmdscale function in R ([www.R-project.org](http://www.R-project.org)) with distances calculated by the Kimura 2-parameter method (Kimura, 1980).

### **Results**

IC-RT-qPCR detected PVY in all plots (Table 1). In plot A, five plants were found to be infected (33.4%) with Ct values in the 25.9 to 33.2 range and melting temperatures between 76.3 and 76.5°C. Plot B exhibited the highest incidence

between foliage samples (80%) with Ct values as low as 26.4 and an average T<sub>m</sub> of 76.7°C ± 0.4 in agreement with the melting temperatures measured in plot A. Only two leaf samples tested positive in plot C (13.4%) with Ct values of 30.1 and 31.4; T<sub>m</sub> values were 76.8 and 78.1, respectively. With respect to tuber samples, they showed a PVY incidence of 86.7% with Ct values as low as 24.3 and an average melting temperature of 77.3°C ± 0.9. An overview of all the T<sub>m</sub> data reveals a 1.8°C difference between the extreme values (76.3°C and 78.1°C), suggesting the presence of sequence variants.

To get a more detailed characterization of the PVY strains infecting *S. tuberosum* cv. Diacol-Capiro in La Unión, a High-throughput transcriptome sequencing was carried out on both foliage and tuber samples; PVY infection was confirmed by ELISA prior sequencing. A total of 4,754,725 paired-end reads were obtained from the infected leaf-sample with 7750 reads per million (RPM) attributed to PVY. Assembly resulted in a consensus PVY genome (LaUnionF) of 9702 nt (excluding the poly-A tail) with an average sequence depth of 3110x (Fig. 1A). The consensus sequence contains 449 polymorphic sites with a transition/transversion ratio of 0.85; the high proportion of polymorphisms suggests the presence of mixed variants within the sample. The ORF encoding the polyprotein is located at positions 188 and 9373; frameshift product P3N-PIPO is predicted at nucleotide positions 2474-2926. The polyprotein ORF contains 399 polymorphic sites that translate into 67 amino acid changes; as expected, most substitutions map to the third codon position (316). Percentagewise, most changes are amino acid changes are located in the 6K1 (7/52, 13.5%), 6K2 protein (5/52, 9.6%), P1 (25/285, 8.8%) and P3 (25/364, 6.8%). CP (3/267, 1.1%) and NIb (13/519, 2.5%) showed the lowest degrees of variation. Isolate PVY\_LaUnionF shares 99.8% sequence identity with a similar isolate from the municipality of Yarumal in the north of Antioquia with phylogenetic affinity to PVY<sup>NTN</sup> strains (Muñoz *et al.*, 2016).

The high levels of variation observed in the PVY foliage assembly may be indicative of sequences experiencing diversifying selection. To address this point, the ratio (K<sub>a</sub>/K<sub>s</sub>) of non-synonymous substitutions per non-synonymous site (K<sub>a</sub>) to

synonymous substitution per synonymous site ( $K_s$ ) was measured along the polyprotein coding region using a sliding window of 14 codons (Fig 2).  $K_a/K_s$  values could be determined for 1831 sites, 1473 (80.4%) were under strong purifying selection ( $K_a/K_s=0$ ), 355 were under weak purifying selection (19.4 %,  $0 < K_a/K_s < 1$ ) and three sites under either neutral or weak positive selection ( $K_a/K_s > 1$ ). All windows under positive selection were located within the P1 protein and centered at amino acid positions 59 ( $K_a/K_s=1.05$ ), 60 ( $K_a/K_s = 1.0$ ) and 63 ( $K_a/K_s = 1.09$ ), comprising the amino acid sequence (I/V)RTSK(N/S)G(T/A). A region centered at residue 2083 within NIa-Pro showed a  $K_a/K_s$  very close to the threshold value (0.97) and comprises a region with the sequence (H/D)LFKSLNGS(T/M)E(A/V).

The tuber transcriptome, on the other hand, resulted in a data set comprising 4,754,725 paired-end reads of which 8256 (1736 RPM) were attributed to PVY. The assembled genome consists of a 9704 nt contig (excluding the polyA tail) with average sequence depth of 170x and was named LaUnionT (Fig. 1B). In contrast to the leaf isolate, the tuber PVY consensus contained only six polymorphic nucleotide sites; two of them, U4746G and T6875C, result in amino acid changes in CI (V1520G) and NIa-Pro (W2230R). The closest PVY isolate available in GenBank corresponds to isolate mar7 infecting *S. lycopersicum* at the municipality of El Peñol in the vicinity of La Unión (Muñoz-Baena *et al.*, 2016). PVY\_mar7 has phylogenetic affinity to the PVY<sup>N</sup> group and shares 99.4% nucleotide sequence identity with PVY\_tuber. Pairwise comparison among the genomes from foliage and tubers reveal significant differences in nucleotide and amino acid sequence identity between regions encoding each of the mature proteins (Table 2). The highest differences were found at the 6K2 and P3 proteins with nucleotide (amino acid) percent identities of 85.9 (94.0) and 92.2 (92.4); the best conserved segment correspond to CP (95.6/99.2) and NIb (95.1/98.5). 5' and 3' untranslated regions share 87.5% and 86.6% nucleotide sequence identity, respectively.

Phylogenetic analysis of complete PVY sequences confirmed that LaUnionF and LaUnionT cluster in two different clades within the PVY<sup>N/NTN</sup> group (Fig. 3A).

PVY\_LaUnionF is part of a group comprising only PVY<sup>NTN</sup> recombinant strains and its close evolutionary relationship with the PVY isolate from Yarumal becomes clear as they cluster together within the clade. The tuber strain, on the other hand, belongs to a group comprising PVY<sup>N</sup> strains and other PVY<sup>NTN</sup> strains; as expected, isolate mar7 from tomato also belongs to this group. The recombinant nature of the PVY\_LaUnionF was further confirmed by recombination analysis (Fig. 3B). Sequence comparison with PVY<sup>N</sup>, PVY<sup>O</sup> and PNY<sup>C</sup> strains reveal three recombination breakpoints at the HC-Pro/P3 and 6K2/Nla-VPg junctions and near the end of the CP region, a pattern typical of PVY<sup>NTN</sup> strains (Karasev & Gray, 2013). Recombination analysis of the PVY tuber strain failed to reveal recombination breakpoints (Fig. 3C).

To corroborate the diversity of PVY strains infecting *S. tuberosum* (cv. Diacol-Capiro) in La Unión, partial sequencing of the CP region was performed in leaf samples from different commercial cultivation plots (LU1-6). Multidimensional scaling analysis using sequences from Colombia and the rest of the world reveals seven distinct PVY clusters (Fig. 4). Cluster A includes sequences belonging to strains N, N:O, W, C and NP from all over the world; Cluster C includes sequences corresponding to strains PVY<sup>N</sup> and PVY<sup>NTN</sup>; the remaining clusters are composed exclusively from Colombian isolates. The CP segment from LaUnionF clustered in group C, which includes the PVY genome from Yarumal, in agreement with phylogenetic analysis discussed above. This group also includes Sanger sequences LU1, LU2, LU4, LU5 and LU6 and Colombia isolates 13Col, 14Col, 40MY and 454749MY previously deposited in GenBank by Gil *et al.* (2011) and Henao *et al.* (2013). Group C comprises only PVY strains with *S. tuberosum* as isolation host. The CP segment LaUnionT clustered in group B composed exclusively by Colombian isolates and includes isolates mar7 (Muñoz-Baena *et al.*, 2016) and 44Col (Jaramillo *et al.*, 2011) infecting tomato and tamarillo (*Solanum betaceum*) respectively; none of the Sanger sequences obtained here was included within this group. Interestingly, sequence LU3 clustered in group A

suggesting the presence of at least a third PVY strain in La Unión. No sequences related to groups C-F were found in this study.

## Discussion

In this work, we have demonstrated the presence of at least two PVY strains infecting *S. tuberosum* (cv. Diacol-Capiro) in the municipality of La Unión, the main potato producer of Antioquia. The isolate PVY\_LaUnionT, infecting tuber-seed material has phylogenetic affinity with strains of the tobacco vein necrosis PVY group, PVY<sup>N</sup> (Singh *et al.*, 2008). PVY<sup>N</sup> isolates were reported for the first time in Peru and Bolivia in the 1940s (Nobrega and Silberschmidt, 1944) followed by reports in the United States and Europe in the 1960s (Kahn and Monroe, 1963). Symptoms caused by PVY<sup>N</sup> are not considered to be severe on potato (Singh *et al.*, 2008; Kerlan and Gray, 2013). Interestingly, Colombian PVY<sup>N</sup> isolates form a very distinct cluster with widespread distribution in Antioquia and the rest of the country (Gil *et al.*, 2011; Henao *et al.*, 2013). Isolates from this group have also been collected in the provinces of Nariño (17BM, 34MY, and 35MY), Boyacá (17BM) and Cundinamarca (13FB). In Antioquia, members of this group have been collected in the municipalities of Santa Rosa de Osos (6MED), La Ceja (38MR) and Sonsón (5MED), the latter two located in the vicinity of La Unión. Group C seems to be able to cross the host barrier as attested by isolates 44Col and mar7 infecting *S. betaceum* and *S. lycopersicum*, respectively (Jaramillo *et al.*, 2011; Muñoz-Baena *et al.*, 2016). PVY\_LaUnionT is very closely related to isolate mar7, which together with the geographical proximity of the municipality of El Peñol, suggests that passage of PVY between alternate hosts might be a result of inadequate crop management practices. The infection of PVY<sup>NTN</sup> and PVY<sup>N</sup> potato-type strains has been previously reported in Spain and Poland (Aramburu *et al.*, 2006; Hasiów-Jaroszewska *et al.*, 2015).

The PVY genome from foliage, LaUnionF, is clearly a member of the recombinant PVY<sup>NTN</sup> group of strains and its similarity to a previously sequenced genome from Yarumal in north Antioquia indicate that both isolates probably have a common origin. Analysis of the CP region shows that PVY\_LaUnionF is part of a group of

PVY strains with worldwide distribution and previously reported in the east of Antioquia in the municipalities of Carmen de Viboral (454749MY), Sonsón (13Col and 14Col) and La Ceja (40MY) (Gil *et al.*, 2011; Henao *et al.*, 2013). The majority of Sanger sequences (LU1-2 and LU4-6) clustered in within this group thus confirming the high prevalence of this strain in the potato cultivating regions of eastern Antioquia. Our data indicates that isolate PVY\_LaUnionF has high sequence variability and comprises a larger set of sequence variants that PVY\_Yarumal, in which only fifteen nucleotide polymorphic sites that translate into only three of amino acid changes in P1 (K30E), 6K1 (M1121I) and NIb (K2758R) (Muñoz *et al.*, 2016). The K30E was also found in the PVY\_LaUnionF and a M1121V change was observed instead of M1121I; no polymorphism equivalent to K2758R was detected in PVY\_LaUnionF.

Sliding window comparison on non-synonymous to synonymous substitution is on good agreement with previous reports that suggest that P1, P3, 6K1, 6K2, and CP genes have more neutral codons than expected (Cuevas *et al.*, 2012). According to our data, P1 is undergoing moderate positive selection in the segment comprising residues 59 to 66. There are few analysis of positive selection on PVY, however, a recent population analysis of PVY from Iran identified 22 sites of P1 and three in NIa-Pro under positive selection (Pourrahim and Farzadfar, 2016). Protein P1 is a serine protease involved in RNA silencing and it has been suggested that modulation of P1 activity has evolved to keep viral amplification below host-detrimental levels therefore helping to maintain a higher long-term replicative capacity (Pasin *et al.*, 2014). It would be interesting if future studies could determine that slight positive selection observed at the N-terminus of P1 in PVY\_LaUnionF is related to fine-tuning of its protease function. The other region undergoing relaxed selection pressure is located at the N-terminus of a second protease: NIa-Pro, which is required for the proteolytic processing of most potyviral proteins and its function, has an important effect on infectivity (Ivanov *et al.*, 2014).

Interestingly, sequence analysis of sample LU3 reveals that this isolate is most similar to isolates CO1960 (99.7 %), CO1827 (99.7%), CO1801 (99.5%) from the

United States (Karasev *et al.*, 2011). These isolates are part of a PVY<sup>O</sup> group of strains reported to be spreading across the United States and other countries world-wide and further classified as a distinct serological group (PVY<sup>O5</sup>), a fact confirmed by comparison of complete genomes. The PVY<sup>O</sup> group can cause mild mosaic in tobacco and induces necrosis and severe stunting in potato cultivars carrying the *Ny* gene. This is the first report in Colombia for this group.

In 2014, Colombia has an estimated yield of 19.9 t ha<sup>-1</sup> which falls short to the productivity seen in countries such as UK (33.17 t ha<sup>-1</sup>) and the United States (51 t ha<sup>-1</sup>) (<http://faostat3.fao.org>). It is likely that low yields are related to the high incidence of viral infections as a result of the lack of adequate seed-certification programs and malpractices in the handling of crops. One of the principal means of virus distribution is through movement of infected propagation material and once a pathogen gets introduced into a new region it may spread rapidly through the action of vectors and cultural malpractices. PVY is a constant threat to potato production as a consequence of the lack of efficient resistance of local varieties to many of its strains and, for this reason the main control strategy is based on the certification of tuber seed production. The high incidence levels of PVY detected in foliage and tuber-seed samples and the finding of at least three different strains in a single potato-producing municipality in Antioquia reveal a worrying situation regarding the sanitary conditions of this crop. Urgent measures are required to strengthen current seed-certification programs such as the incorporation of methods such as IC-RT-qPCR and a constant monitoring of the genomic characteristics of the PVY strains present in each potato-producing region. The latter is of the utmost importance as genome information can help to predict the epidemiological impact of PVY infection, since it is possible to associate the strain genotypes with their impact on yield and tuber quality as well as with alternate hosts and transmission efficiency by arthropod vectors.

### **Acknowledgment**

This work was funded by Universidad Nacional de Colombia (Grant VRI: 34585) and International Foundation for Science (Sweden, Grant C/4634-2).

## Literature cited

Aramburu, J., L. Galipienso and M. Matas. 2006. Characterization of *Potato virus Y* isolates from tomato crops in northeast Spain. *Eur. J. Plant Pathol.* 115(2), 247-258. Doi: 10.1007/s10658-006-9003-x.

Blanco-Urgoiti, B., F. Sánchez, J. Dopazo and F. Ponz. 1996. A strain-type clustering of *Potato virus Y* based on the genetic distance between isolates calculated by RFLP analysis of the amplified coat protein gene. *Arch. Virol.* 141(12), 2425-42. Doi: 10.1007/BF01718641.

Cuevas, J.M., A. Delaunay, M. Rugar, E. Jacquot and S.F. Elena. 2012. Molecular evolution and phylogeography of *Potato virus Y* based on the CP gene. *J. Gen. Virol.* 93(11), 2496-2501. Doi: 10.1099/vir.0.044347-0.

Fageria, M.S., M. Singh, U. Nanayakkara, Y. Pelletier, X. Nie and D. Wattie. 2013. Monitoring current-season spread of *Potato virus Y* in potato fields using ELISA and real-time RT-PCR. *Plant Dis.* 97(5), 641-644. Doi: 10.1094/PDIS-03-12-0283-RE.

Gallo, Y., P. Gutiérrez and M. Marín. 2012. Generación de anticuerpos policlonales para la detección de la variante genotípica GIII de PVY, en cultivos de tomate de árbol y papa de Colombia. *Rev. Col. Biotecnol.* 14(1): 245-255.

Gil, J.F., J.M. Cotes and M. Marín. 2011. Incidencia de potyvirus y caracterización molecular de PVY en regiones productoras de papa (*Solanum tuberosum* L.) de Colombia. *Rev. Col. Biotecnol.* 13(1), 85-93.

Gish, W. and D.J. States. 1993. Identification of protein coding regions by database similarity search. *Nature Genet.* 3(3), 266-272. Doi: 10.1038/ng0393-266.

Glais, L., M. Tribodet and C. Kerlan C. 2002. Genomic variability in *Potato potyvirus Y* (PVY): evidence that PVY<sup>NW</sup> and PVY<sup>NTN</sup> variants are single to multiple recombinants between PVY<sup>O</sup> and PVY<sup>N</sup> isolates. *Arch. Virol.* 147(2), 363-378. Doi: 10.1007/s705-002-8325-0.

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman and A. Regev. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29(7), 644-652. Doi: 10.1038/nbt.1883.

Hasiów-Jaroszewska, B., J. Stachecka, J. Minicka, M. Sowiński and N. Borodynko. 2015. Variability of *Potato virus Y* in tomato crops in Poland and development of a reverse-transcription loop-mediated isothermal amplification method for virus detection. *Phytopathology.* 105(9), 1270-1276. Doi: 10.1094/PHYTO-08-14-0219-R.

Henao E., P. Gutiérrez and M. Marín. 2013. Análisis filogenético de aislamientos del *Potato virus Y* (PVY) obtenidos en cultivos de papa (*Solanum tuberosum*) y tomate de árbol (*Solanum betaceum*) en Colombia. *Actual. Biol.* 35(99), 219-232.

Ivanov, K.I., K. Eskelin, A. Lõhmus and K. Mäkinen. 2014. Molecular and cellular mechanisms underlying potyvirus infection. *J. Gen. Virol.* 95(7), 1415-1429. Doi:10.1099/vir.0.064220-0.

Jaramillo, M., P. Gutiérrez, L.E. Lagos, J.M. Cotes and M. Marín. 2011. Detection of a complex of viruses in tamarillo (*Solanum betaceum*) orchards in the Andean region of Colombia. *Trop. Plant Pathol.* 36(3), 150-159. Doi: 10.1590/S1982-56762011000300003.

Kahn, R.P. and R.I. Monroe. 1963. Detection of tobacco veinal necrosis (strain of *Potato virus Y*) in *Solanum carensaii* and *S. andigenum* introduced into the United States. *Phytopathology.* 53, 1356-1359.

Karasev, A.V., X. Hu, C.J. Brown, C. Kerlan, O.V. Nikolaeva, J.M. Crosslin and S.M. Gray. 2011. Genetic diversity of the ordinary strain of *Potato virus Y* (PVY) and origin of recombinant PVY strains. *Phytopathology.* 101(7), 778-785. Doi: 10.1094/PHYTO-10-10-0284.

Karasev, A.V. and S.M. Gray. 2013. Continuous and emerging challenges of *Potato virus Y* in potato. *Annu. Rev. Phytopathol.* 51(1), 571-586. Doi: 10.1146/annurev-phyto-082712-102332.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16(2), 111-120.

Langmead, B. and S. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9(4), 357-359. Doi: 10.1038/nmeth.1923.

Medina, H.C., P. Gutiérrez and M. Marín. 2015. Detección del *Potato virus Y* (PVY) en tubérculos de papa mediante TAS-ELISA y qRT-PCR en Antioquia (Colombia). *Bioagro.*27(2), 83-92.

Milne, I., M. Bayer, L. Cardle, P. Shaw, G. Stephen, F. Wright, and D. Marshall. 2010. Tablet-next generation sequence assembly visualization. *Bioinformatics.* 26(3), 401–402. Doi: 10.1093/bioinformatics/btp666

Muñoz-Baena, L., P. Gutiérrez and Marín-Montoya. 2016. Detección y secuenciación del *Potato virus Y* (PVY) que infecta plantas de tomate en Antioquia (Colombia). *Bioagro.* 28(2), 69-80.

Muñoz, D., P. Gutiérrez and M. Marín. 2016. Detección y caracterización molecular del *Potato virus Y* (PVY) en cultivos de papa (*Solanum tuberosum* L.) del norte de Antioquia, Colombia. *Rev. Prot. Veg.* 31(1), 1-11.

Nei, M. and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3(5), 418-426.

Nobrega, N.R. and K. Silberschmidt. 1944. Sobre una provavel variante do virus « Y » da batatinha que tem a peculiaridade de provocar necroses em plantas de fumo. *Arq. Inst. Biol.* 1944, 307-330.

Nolte, P., J.L. Whitworth, M.K. Thornton and C.S. McIntosh. 2004. Effect of seedborne *Potato virus Y* on performance of Russet Burbank, Russet Norkotah, and Shepody potato. *Plant Dis.* 88(3), 248-252. Doi: 10.1094/PDIS.2004.88.3.248.

Ogawa, T., Y. Tomitaka, A. Nakagawa and K. Ohshima. 2008. Genetic structure of a population of *Potato virus Y* inducing potato tuber necrotic ringspot disease in Japan; comparison with North American and European populations. *Virus Res.* 131(2), 199-212. Doi: 10.1016/j.virusres.2007.09.010.

Olsper, A., B.Y. Chung, J.F. Atkins, J.P. Carr and A.E. Firth. 2015. Transcriptional slippage in the positive-sense RNA virus family *Potyviridae*. *EMBO Reports.* 16(8), 995-1004. Doi: 10.15252/embr.201540509.

Pasin, F., C. Simón-Mateo and J.A. García. 2014. The hypervariable amino-terminus of P1 protease modulates potyviral replication and host defense responses. *PLoS Pathog.* 10(3), e1003985. Doi: 10.1371/journal.ppat.1003985.

Pourrahim, R. and S. Farzadfar. 2016. Population analysis of Iranian *Potato virus Y* Isolates using complete genome sequence. *Plant Pathol. J.* 32(1), 33-46. Doi: 10.5423/PPJ.OA.07.2015.0144.

Quenouille, J., N. Vassilakosa and B. Moury. 2013. *Potato virus Y*: a major crop pathogen that has provided major insights into the evolution of viral pathogenicity. *Mol. Plant Pathol.* 14(5), 439-452. Doi: 10.1111/mpp.12024.

Revers, F. and J.A. García. 2015. Molecular biology of potyviruses. *Adv. Virus Res.* 92:101-199. Doi: 10.1016/bs.aivir.2014.11.006.

Schena, L., F. Nigro, A. Ippolito and D. Gallitelli. 2004. Real-time quantitative PCR: a new technology to detect and study phytopathogenic and antagonistic fungi. *Eur. J. Plant Pathol.* 110(9), 893-908. Doi: 10.1007/s10658-004-4842-9.

Scholthof, K.B., S. Adkins, H. Czosnek, P. Palukaitis, E. Jacquot, T. Hohn, B. Hohn, K. Saunders, T. Candresse, P. Ahlquist, C. Hemenway and G.D. Foster. 2011. Top 10 plant viruses in molecular plant pathology. *Mol. Plant Pathol.* 12(9), 938-954. Doi:10.1111/j.1364-3703.2011.00752.x.

Singh, R.P., J.P.T. Valkonen, S.M. Gray, N. Boonham, R.A.C. Jones, C. Kerlan and J. Schubert. 2008. Brief review: The naming of *Potato virus Y* strains infecting potato. Arch. Virol., 153(1), 1-13. Doi: 10.1007/s00705-007-1059-1.

Singh, M., R.P. Singh, M.S. Fageria, X. Nie, R. Coffin and G. Hawkins. 2013. Optimization of a Real-Time RT-PCR assay and its comparison with ELISA, conventional RT-PCR and the grow-out test for large scale diagnosis of *Potato virus Y* in dormant potato tubers. Am. J. Potato Res. 90(1), 43-50. Doi: 10.1007/s12230-012-9274-z.

Tamura, K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. Mol. Biol. Evol. 9(4), 678-687.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar. 2013. MEGA6: Molecular evolutionary genetics analysis. Mol. Biol. Evol. 30(12), 2725-2729. Doi: 10.1093/molbev/mst197.

Untiveros, M., A. Olsper, K. Artola, A.E. Firth, J.F. Kreuze and J.P.T. Valkonen. 2016. A novel sweet potato potyvirus open reading frame (ORF) is expressed via polymerase slippage and suppresses RNA silencing. Mol. Plant Pathol. doi:10.1111/mpp.12366.

Wetzel, T., T. Candresse, G. Macquaire, M. Ravelonandro and J. Dunez. 1992. A highly sensitive immunocapture polymerase chain reaction method for plum pox potyvirus detection. J. Virol. Methods. 39(1-2), 27-37. Doi: 10.1016/0166-0934(92)90122-T.

**Table 1.** IC-RT-qPCR detection of PVY in leaf samples from three potato cultivating plots (A-C) and tubers (T) from La Unión (Antioquia).

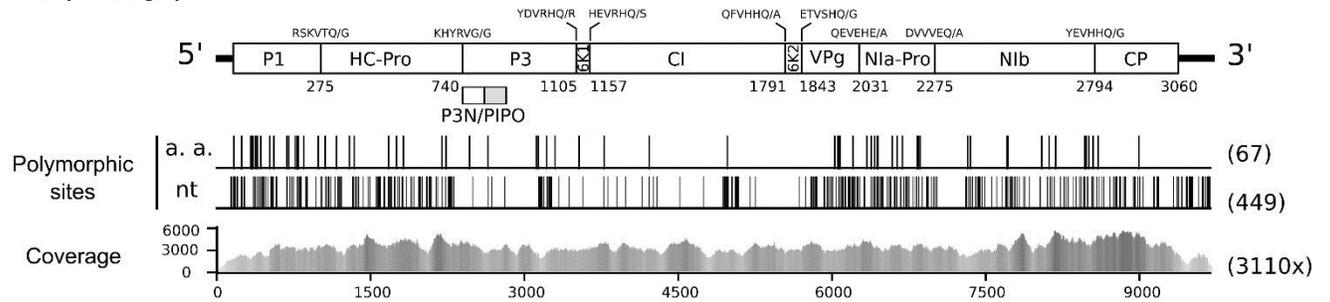
	Plot A		Plot B		Plot C		Tubers				
	Ct*	Tm** (°C)	Ct	Tm (°C)	Ct	Tm (°C)	Ct	Tm (°C)			
<b>A1</b>	>35	-	<b>B1</b>	>35	-	<b>C1</b>	>35	-	<b>T1</b>	28.1	77.8
<b>A2</b>	>35	-	<b>B2</b>	29.8	77.2	<b>C2</b>	>35	-	<b>T2</b>	27.1	77.2
<b>A3</b>	>35	-	<b>B3</b>	26.4	76.4	<b>C3</b>	>35	-	<b>T3</b>	30.9	76.6
<b>A4</b>	30.3	76.5	<b>B4</b>	29.1	77.1	<b>C4</b>	>35	-	<b>T4</b>	28.2	78.1
<b>A5</b>	>35	-	<b>B5</b>	29.2	77.2	<b>C5</b>	>35	-	<b>T5</b>	27.4	77.4
<b>A6</b>	28.7	76.5	<b>B6</b>	30.4	77.1	<b>C6</b>	>35	-	<b>T6</b>	27.2	77.6
<b>A7</b>	33.2	76.3	<b>B7</b>	>35	-	<b>C7</b>	>35	-	<b>T7</b>	30.1	77.5
<b>A8</b>	>35	-	<b>B8</b>	28.8	76.4	<b>C8</b>	>35	-	<b>T8</b>	>35	-
<b>A9</b>	>35	-	<b>B9</b>	28.7	76.3	<b>C9</b>	>35	-	<b>T9</b>	32.1	78.1
<b>A10</b>	>35	-	<b>B10</b>	28.3	76.9	<b>C10</b>	30.1	78.1	<b>T10</b>	31.1	77.5
<b>A11</b>	28.2	76.5	<b>B11</b>	29.6	76.3	<b>C11</b>	>35	-	<b>T11</b>	>35	-
<b>A12</b>	>35	-	<b>B12</b>	28.2	76.5	<b>C12</b>	31.4	76.8	<b>T12</b>	>35	-
<b>A13</b>	>35	-	<b>B13</b>	>35	-	<b>C13</b>	>35	-	<b>T13</b>	24.3	77.4
<b>A14</b>	25.9	76.3	<b>B14</b>	29.4	76.56	<b>C14</b>	>35	-	<b>T14</b>	23.7	78.1
<b>A15</b>	>35	-	<b>B15</b>	29.6	76.86	<b>C15</b>	>35	-	<b>T15</b>	25.7	77.6

\* Ct: Threshold cycles below 35 were considered positive. Tm: melting temperatures.

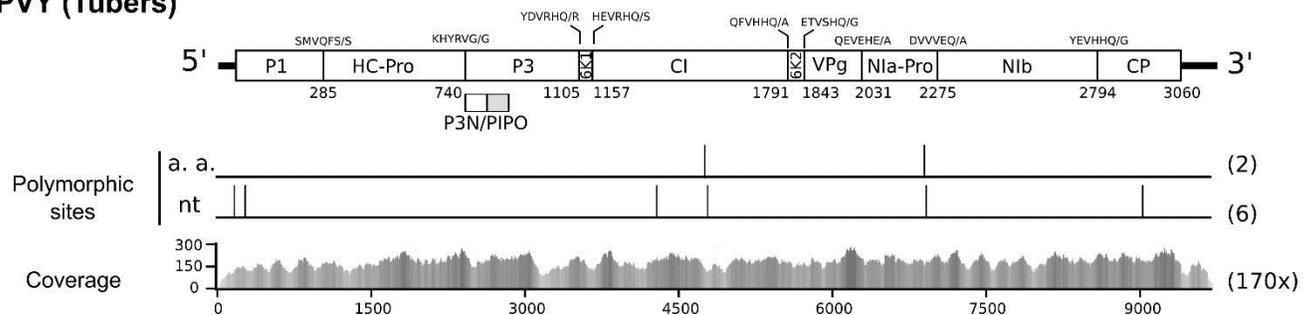
**Table 2.** Pairwise sequence comparison the PVY genomes infecting tubers and foliage in La Unión (Antioquia). Values correspond to the percent identity between the consensus sequences at the nucleotide (nt) and amino acid level (a.a.).

		5'UTR	P1	HC-Pro	P3	6K1	CI	6K2	Nla-VPg	Nla-Pro	Nib	CP	3'UTR
% id	(nt)	87.5	92.2	93.1	85.9	82.7	85.4	84.1	91.9	94.4	95.1	95.6	86.5
	(a.a)	-	92.4	97.6	94.0	86.5	95.4	90.4	96.8	97.5	98.5	99.2	-

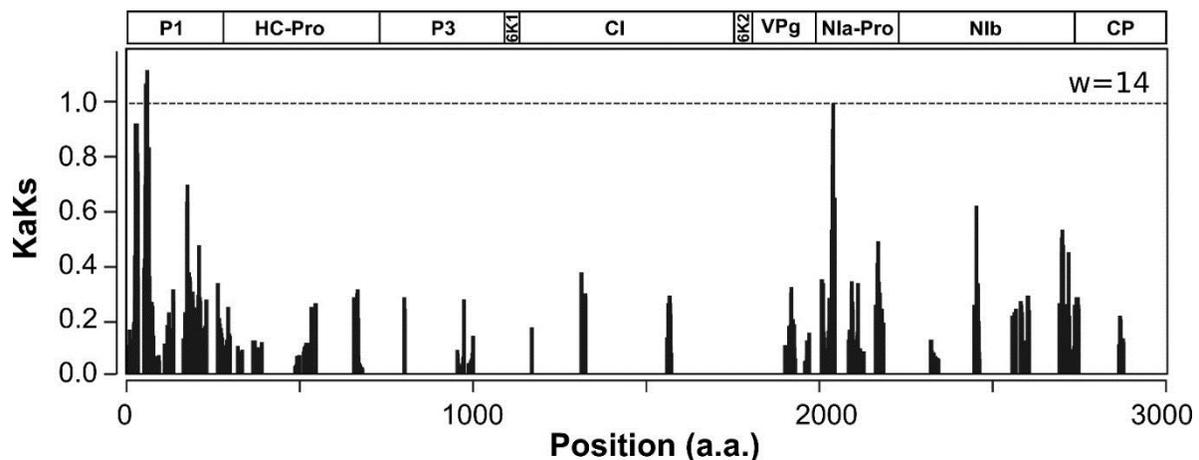
### PVY (Foliage)



### PVY (Tubers)



**Figure 1.** Sequence coverage and mapping of the PVY genomes infecting potato foliage and tubers in La Unión (Antioquia). The relative position of each mature protein in PVY is shown on top of each panel; viral protease cleavage sites are shown. The PVY isolated from foliage (PVY\_LaUnionF) exhibits a very high sequence coverage (3110 x) and variability with 449 polymorphic sites suggesting the presence of mixed variants in the sample. In contrast, the PVY strain from foliage (PVY\_LaUnionT), resulted from an assembly with 170x coverage and only six polymorphisms.

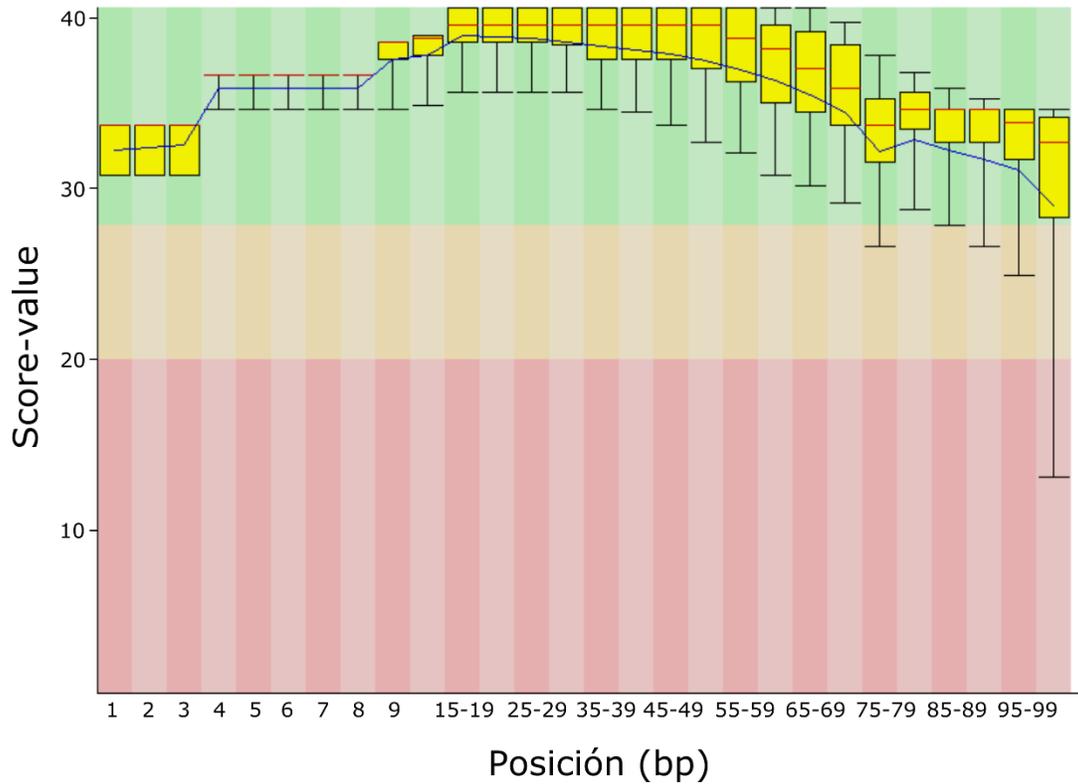




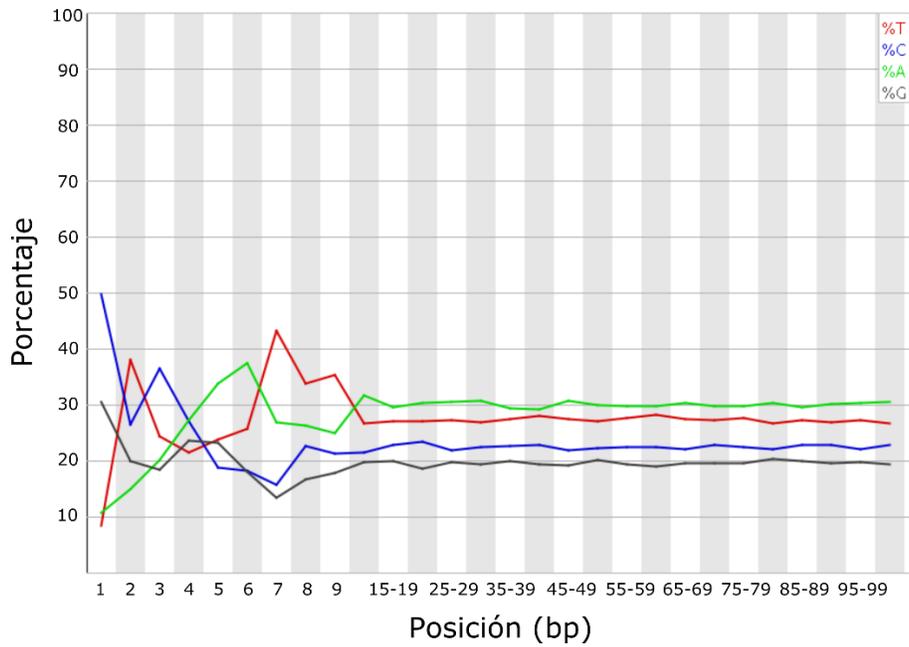


### 4.3.1. ANEXOS – Capítulo 3

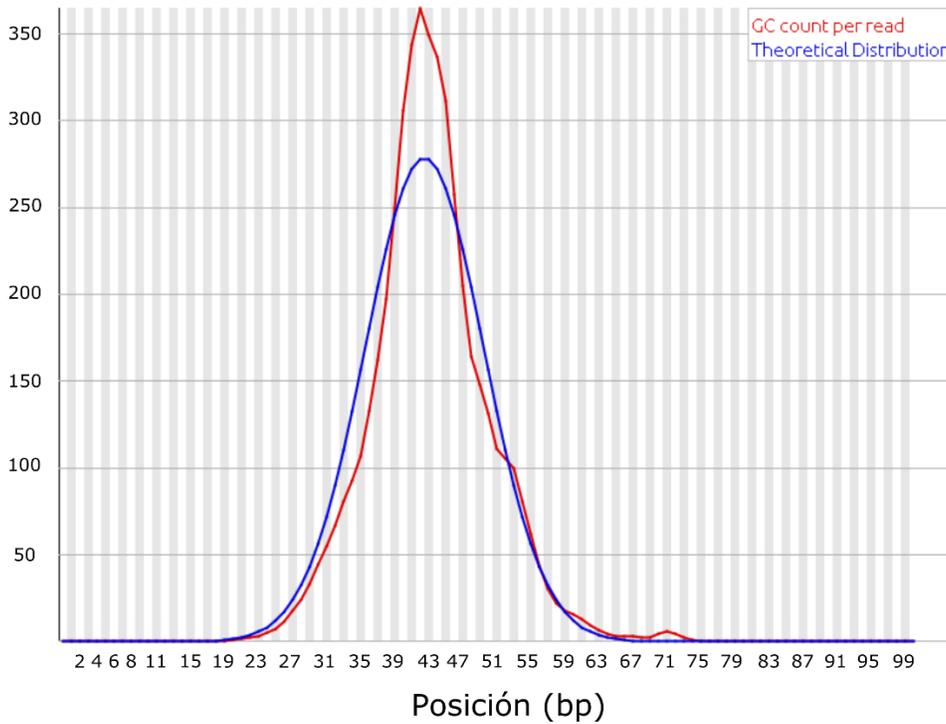
#### 4.3.1.1. Análisis de los datos provenientes de la secuenciación HiSeq 2000 (Illumina)



**Anexo 1.** Los valores de calidad de los *reads* se representan mediante un diagrama Box-Whisker. La línea azul indica la calidad promedio. El fondo de la gráfica divide el eje de las ordenadas en tres zonas: buena calidad (verde), calidad razonable (naranja) y mala calidad (roja). En la mayoría de sistemas de secuenciación, es común observar un decaimiento hacia el final de las secuencias.



**Anexo 2.** Proporción de los nucleótidos a lo largo de la longitud de los *reads*.



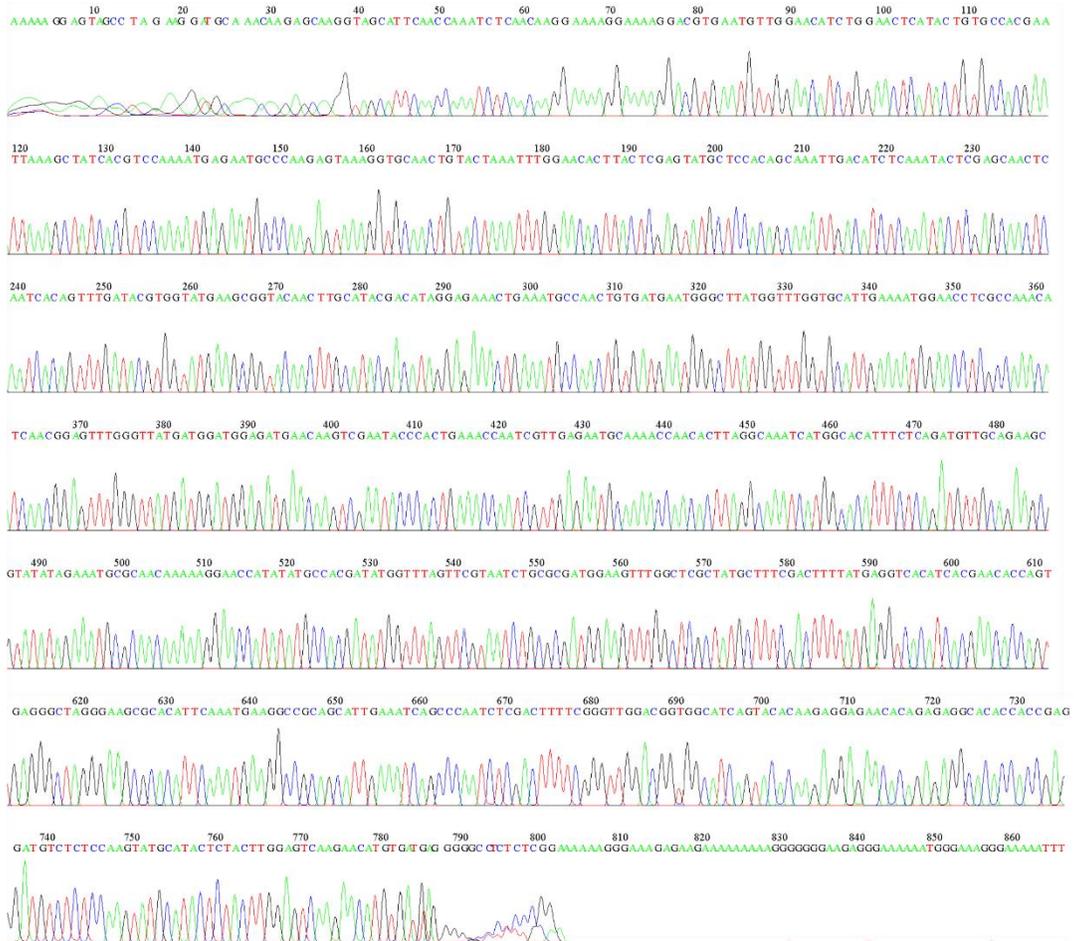
**Anexo 3.** Distribución del contenido de guanina-citosina (GC) a lo largo de toda la secuencia (rojo) en comparación con una distribución normal (azul). El pico central corresponde al contenido GC del genoma.



File: 62mar\_PVYCPF.ab1  
Sample: 62mar\_PVYCPF

Run Ended: 2015/4/3 23:11:36  
Lane: 20

Signal G:2673 A:3311 C:2888 T:2593  
Base spacing: 15.245707  
892 bases in 10791 scans



Anexo 5. Secuenciación Sanger para la muestra LU2.

Description	Max score	Total score	Query cover	E value	Ident	Accession
Potato virus Y isolate Yarumal_varB, complete genome	1366	1366	100%	0.0	99%	KT336552.1
Potato virus Y isolate Yarumal, complete genome	1366	1366	100%	0.0	99%	KT336551.1
Potato virus Y isolate La Union polyprotein gene, complete cds	1366	1366	100%	0.0	99%	KR149260.1
Potato virus Y isolate FI:50 7a:06 polyprotein gene, partial cds	1361	1361	100%	0.0	99%	JX432987.1
Potato virus Y isolate FI:413 2b:06 polyprotein gene, partial cds	1361	1361	100%	0.0	99%	JX432982.1
Potato virus Y isolate FI:V213 16:07 polyprotein gene, partial cds	1361	1361	100%	0.0	99%	JX432978.1
Potato virus Y isolate FI:V3 17:07 polyprotein gene, partial cds	1361	1361	100%	0.0	99%	JX432976.1
Potato virus Y isolate FI:V12 2:07 polyprotein gene, partial cds	1361	1361	100%	0.0	99%	JX432974.1
Potato virus Y isolate FI:V322 12:07 polyprotein gene, partial cds	1361	1361	100%	0.0	99%	JX432971.1
Potato virus Y isolate 3D, complete genome	1361	1361	100%	0.0	99%	KJ946936.2
Potato virus Y strain NTN, complete genome	1361	1361	100%	0.0	99%	KM396648.1
Potato virus Y isolate LUNG 4, complete genome	1361	1361	100%	0.0	99%	JF927752.1
Potato virus Y isolate La_Union_1 capsid protein gene, partial cds	1359	1359	97%	0.0	100%	KR902698.1
Potato virus Y isolate F17, complete genome	1357	1357	100%	0.0	99%	KX184817.1
Potato virus Y isolate FI:PAAVO:07 polyprotein gene, partial cds	1357	1357	100%	0.0	99%	JX432992.1
Potato virus Y isolate FI:MTTJOK3:07 polyprotein gene, partial cds	1357	1357	100%	0.0	99%	JX432991.1
Potato virus Y isolate FI:308 9b:06 polyprotein gene, partial cds	1357	1357	100%	0.0	99%	JX432990.1
Potato virus Y isolate FI:180 6b:06 polyprotein gene, partial cds	1357	1357	100%	0.0	99%	JX432989.1
Potato virus Y isolate FI:202 8c:06 polyprotein gene, partial cds	1357	1357	100%	0.0	99%	JX432983.1
Potato virus Y isolate FI:449 4b:06 polyprotein gene, partial cds	1357	1357	100%	0.0	99%	JX432980.1
Potato virus Y isolate FI:V261 3:07 polyprotein gene, partial cds	1357	1357	100%	0.0	99%	JX432979.1
Potato virus Y isolate FI:V65 3:07 polyprotein gene, partial cds	1357	1357	100%	0.0	99%	JX432975.1
Potato virus Y isolate FI:V252 19:07 polyprotein gene, partial cds	1357	1357	100%	0.0	99%	JX432973.1
Potato virus Y isolate FI:V121 10:07 polyprotein gene, partial cds	1357	1357	100%	0.0	99%	JX432972.1

**Anexo 6.** Resultado de BLASTN usando como *query* las secuencias confirmadas por Sanger (LU1).

## 5. ANEXOS GENERALES

### 5.1. Análisis bioinformático para el ensamble de virus mediante datos de secuenciación por NGS

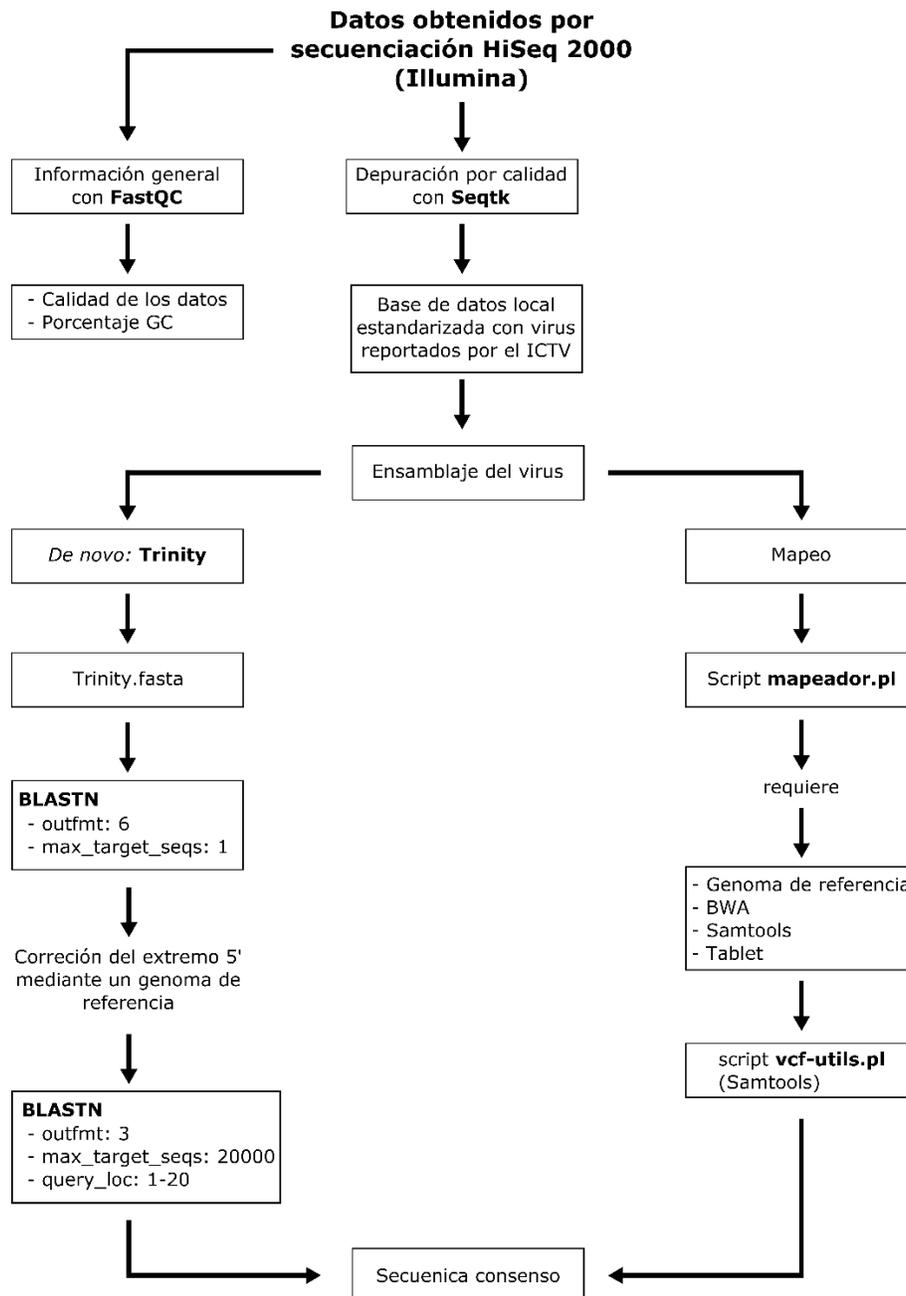


Diagrama del proceso general para el ensamble de PVY mediante datos de NGS.

## 5.2. Script *mapeador.pl* (realizado en colaboración con Mateo Duque Villegas, MSc.)

```
#!/usr/bin/perl

use strict;
use warnings;

# El programa requiere tres argumentos de entrada ingresado
por el usuario para poder iniciar.

my $num_args = $#ARGV + 1;
if ($num_args != 3) {
    print "\nIngrese los tres elementos necesarios para correr
el programa con su respectiva ubicación. Ejemplo: \nperl
mapeador.pl /home/path/to/file/genoma_ref.fasta
/home/path/to/file/libreria1.fastq.gz
/home/mateo/unal/libreria2.fastq.gz\n\n ";
    exit;
}

#Se crea una carpeta con archivos temporales con el nombre
outdir_tmp. No lo hace en el evento que ya exista. Al final
se elimina esta carpeta y los archivos innecesarios.

my $outdir_tmp = 'mapeo_out';
unless (-e $outdir_tmp) {
    system("mkdir $outdir_tmp");
}
my $ref= $ARGV[0]; #Genoma de referencia
my $lib1 = $ARGV[1]; #Librería 1 (format fastq.gz)
my $lib2 = $ARGV[2]; #Librería 2 (format fastq.gz)

#Se copia el archivo del genoma de referencia a la carpeta
temporal recién creada, puesto que debe leerse al tiempo con
los índices generados en el primer paso, como se muestra a
continuación:

system("cp $ref $outdir_tmp/");

# 1) Indexar archivo de referencia o chequear si ya fue
indexado

my $r1 = (split(/\//,$ref))[-1];
my $r2 = (split(/\./,$r1))[0];
my $fr = "$outdir_tmp/$r1";
```

```

my$fr1 = "$outdir_tmp/$r1.bwt";
my$fr2 = "$outdir_tmp/$r1.amb";
my$fr3 = "$outdir_tmp/$r1.sa";
my$fr4 = "$outdir_tmp/$r1.ann";
my$fr5 = "$outdir_tmp/$r1.pac";

#Si los archivos ya fueron creados previamente, este comando
evita generarlos de nuevo

if (-e $fr1 and -e $fr2 and -e $fr3 and -e $fr4 and -e $fr5){
    print "La referencia ya fue indexada! Continuando...\n"}
else{
    system("bwa index -a bwtsv $ref -p $fr");
#Llama al programa BWA para indexar el genoma de referencia
}

# 2) Indexar primera librería contra la referencia o chequear
si ya fue indexada

my$l1 = (split(/\.\/,(split(/\/\/,$lib1))[-1]))[0]; #Ignora la
extensión .fastq.gz
my$fn1 = "$outdir_tmp/$l1.sai";
if (-e $fn1) {
    print "El archivo $fn1 ya existe! Continuando...\n";}
else{
    system("bwa aln -t 7 $fr $lib1 > $fn1");
}

# 3) Indexar segunda librería contra la referencia o chequear
si ya fue indexada

my$l2 = (split(/\.\/,(split(/\/\/,$lib2))[-1]))[0];
my$fn2 = "$outdir_tmp/$l2.sai";
if (-e $fn2){
    print "El archivo $fn2 ya existe! Continuando...\n"}
else{
    system("bwa aln -t 7 $fr $lib2 > $fn2");
}

# 4) Hacer el mapeo con los archivos índices y librerías o
chequear si el .sam ya existe

my$fn3 = "$outdir_tmp/aln.sam";
if (-e $fn3){
    print "El archivo $fn3 ya existe! Continuando...\n";}
else{
    system("bwa sampe $fr $fn1 $fn2 $lib1 $lib2 > $fn3");
}

```

```

    }
# 5) Convertir el archivo .sam en comprimido .bam que ocupa
menos memoria

my$fn4 = "$outdir_tmp/aln.bam";
my$fn5 = "aln_${r2}_${l1}";
if (-e $fn4 and -e "$fn5.bam" and -e "$fn5.bai"){
    print "Los archivos necesarios para el alineamiento ya
han sido creados: $fn5.bam\ny $fn5.bai. Continuando...
\n\n";} #De nuevo, verifica si los archivos ya han sido
creados
else{

#Llama al programa Samtools para formatear el archivo .sam

system("samtools view -bS $fn3 > $fn4 ; samtools sort $fn4
$fn5 ; samtools index $fn5.bam $fn5.bai"); #Ejecuta comandos
de samtools y luego elimina la carpeta con todos los archivos
que no se necesitan
    }
system("rm -rf $outdir_tmp"); #Elimina la carpeta temporal de
archivos que no son requeridos

# 6) Generar secuencia consenso con calidad

#Llama la rutina vcfutils.pl del programa Samtools

system ("samtools mpileup -uf $ref $fn5.bam | bcftools view -
cg - | vcfutils.pl vcf2fq > cns_${fn5}.fq");

# 7) Visualizar en Tablet

system("tablet $fn5.bam $ref"); #Tablet require de un archivo
.bam (previamente indexado) y el genoma de referencia para
visualizar el ensamblaje.

```

### 5.3. Identificación de polimorfismos

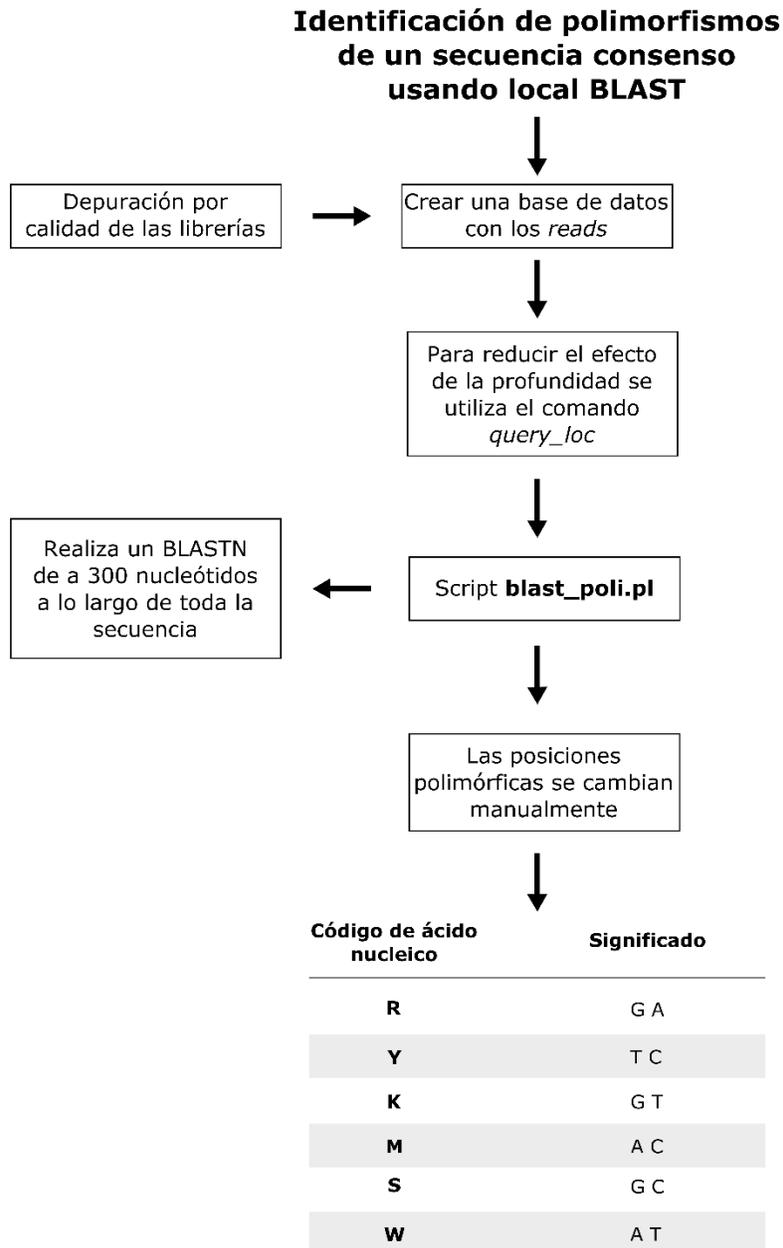


Diagrama para la identificación de polimorfismos en una secuencia consenso mediante local BLAST.

## 6. CONCLUSIONES

En este estudio se ha demostrado la presencia de PVY en zonas de importancia agroeconómica para la producción de papa en Antioquia. En cultivos del norte del departamento se detectó el PVY en un 88.8% de 45 muestras foliares. En el Oriente, dicho porcentaje se encontró entre el 33.4 y 80 %, mientras que en muestras comerciales de tubérculos-semilla, la incidencia del virus fue de un 86.7%. Dichos resultados resaltan la urgencia de revisar los programas de manejo de enfermedades virales en la agroindustria del cultivo de papa en Colombia y, muy especialmente, reforzar los programas de certificación de tubérculo-semilla con el empleo de técnicas de detección altamente sensibles, como las basadas en RT-PCR convencional y RT-PCR en tiempo real. Adicionalmente, se obtuvieron las secuencias completas de tres genomas diferentes de PVY infectando *S. tuberosum* en el norte y oriente de Antioquia, a partir de la Secuenciación de Nueva Generación de muestras de tejido foliar y brotes de tubérculos. Estos genomas comprendieron entre 9701 y 9704 nt, y codifican para una poliproteína de 3060-3061 residuos. Las secuencias de los virus encontrados en el oriente infectando tejido foliar fueron altamente polimórficas, con 399 sitios variables a nivel de nucleótidos y 67 cambios en aminoácidos de la poliproteína, sugiriendo la presencia de mezclas de variantes. En comparación, los genomas de los virus encontrados en tubérculos de esta misma región, mostraron poca variabilidad, pues sólo se observaron seis sustituciones a nivel de nucleótidos y dos cambios a nivel de aminoácidos. Estas diferencias en variabilidad pueden atribuirse a los distintos métodos de transmisión que ocurren en el tejido foliar (por ejemplo, transmisión mecánica, por vectores, tubérculo-semilla), en contraste con el inóculo primario exclusivo del tubérculo-semilla.

En el estudio también se ha demostrado que las simulaciones de dinámica molecular pueden ser utilizadas para estudiar el procesamiento proteolítico de potyvirus mediado por NIa-Pro. Esta información puede usarse para estimar tasas relativas de corte e incorporarse en modelos cinéticos que permitan predecir los

tiempos de maduración de poliproteínas en diferentes potyvirus. La validación del modelo computacional con la información experimental disponible para TEV, sugiere que la dinámica del sitio de corte en el complejo enzima-sustrato es un buen predictor de las tasas relativas de hidrólisis. Sin embargo, las predicciones para PVY deben ser verificadas experimentalmente para comprobar la aplicabilidad general del método desarrollado.