Biblioteca Digital
F C E N - U B A

BIBLIOTECA CENTRAL LUIS F LELOIR
BIBLIOTECA CENTRAL LELOIR
FACULTAD DE CIENCIAS EXACTAS Y NATURALES UBA

Tesis Doctoral

Proteínas virales intrínsecamente desordenadas: disección estructural de la oncoproteína E7 del virus del papiloma humano y de la fosfoproteína P del virus sincicial respiratorio

Noval, María Gabriela

2016-04-19

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:
Noval, María Gabriela. (2016-04-19). Proteínas virales intrínsecamente desordenadas: disección estructural de la oncoproteína E7 del virus del papiloma humano y de la fosfoproteína P del virus sincicial respiratorio. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:
Noval, María Gabriela. "Proteínas virales intrínsecamente desordenadas: disección estructural de la oncoproteína E7 del virus del papiloma humano y de la fosfoproteína P del virus sincicial respiratorio". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2016-04-19.

EXACTAS UBA
Facultad de Ciencias Exactas y Naturales

UBA
Universidad de Buenos Aires

Dirección: Biblioteca Central Dr. Luis F. Leloir, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Intendente Güiraldes 2160 - C1428EGA - Tel. (++54 +11) 4789-9293

Contacto: digital@bl.fcen.uba.ar

UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales

Departamento de Química Biológica

“Proteínas virales intrínsecamente desordenadas: disección estructural de la oncoproteína E7 del virus del papiloma humano y de la fosfoproteína P del virus sincicial respiratorio”

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires en el área Química Biológica

Lic. María Gabriela Noval

Director de tesis: Dr. Gonzalo de Prat Gay

Consejero de estudios: Dr. Julio J. Caramelo

Lugar de trabajo: Laboratorio de Estructura-Función e Ingeniería en Proteínas.
Fundación Instituto Leloir e IIBBA-CONICET, Buenos Aires, Argentina.

Buenos Aires, 2016

Fecha de defensa: 19/04/2016

AGRADECIMIENTOS

En primer lugar quiero agradecer a mi director **Gonzalo**. Por haber confiado en mi siempre, por haberme guiado a lo largo de estos años. Por haberme escuchado cuando más lo necesite, por toda la ayuda y los consejos. Muchas gracias Gonzalo!

A todo el lab 209! El de antes, el de ahora. Por las discusiones, los mates, las charlas, los consejos. En particular quiero agradecerle a Lucia por haberme guiado en mis primeros pasos en el laboratorio. A Mari, Sil, Gaby, Ivi, Leo, Nico, y Sebas por hacer más divertidos mis días en el laboratorio y por la ayuda incondicional de siempre.

A todos los que me ayudaron con la escritura de la tesis, por el aliento y las correcciones. A Lucia, Sebas, Silvina, Mariana y a mi mayor corrector Fran.

A toda la gente de FIL, Dora, a Lili, los chicos de droguero, a Moni y Juan de Biblioteca, a Silvana, a la gente de compras, de administración.

A los miembros de mi comité de seguimiento de tesis (Dr. Ricardo Wolosiuk, Julio Caramelo, Sebastián Klinke), por sus críticas siempre constructivas y por haber ayudado a mejorar mi trabajo.

A todos los amigos que me regalo este instituto, a mis Diosas Leloirenses (Elvis, Ivi, Marie, Juli, Sole Mac, Carol, Ine, Kari, Sofi) por las meriendas, nuestras tardes de danza, la música, las charlas, la amistad incondicional, las quiero bellas! A Sole labanda por ser una de mis dos tucumanas favoritas. A Carol e Ine por tantos momentos compartidos, porque no puedo no adorarlas! A todos mis compañeros de los labos vecinos por la buena onda, por las discusiones sobre clonados, ciencia, por los reactivos prestados, ayuda con equipos, por tantos momentos divertidos.

A Carli por ser como mi hermana. Por quererme a pesar de mis cuelgues, mis desapariciones, por estar siempre, por ser la mejor compañera de viajes. Por ser mi AMIGA.

A la muchachada, mis amigos incondicionales Rox, Agus, Regi, Sil, Tomi, Juli y al nuevo integrante (Il Ruben).

A mis tíos, por estar siempre.

A mi familia adoptiva los Izzos (Gaby, Carlos, y Fede). Por haberme aceptado como una más, por cuidarme, apoyarme, ayudarme y por hacerme sentir siempre tan querida.

A Dieguito, por ser el mejor hermano que la vida me pudo haber dado. Por sus sonrisas que valen más que mil palabras.

A mi papá...por donde empezar. Por TODO. Por el apoyo, la ayuda, la paciencia, por saber que siempre puedo contar con él. Te quiero papá.

A la vida, por haberme regalado 29 años maravillosos con la mejor mamá que podría haberme tocado.

A Fran mi compañero de aventuras. Por TODO. Porque te AMO.

Por haberme compartido tu amor y pasión por la ciencia, por haber sido mi ejemplo en todo, por haberme enseñado que nada es imposible, por haberme apoyado en todas mis locuras, porque siempre fuiste, sos y serás mi mejor amiga…

Esta tesis esta dedicada a vos MAMÁ…

“Proteínas virales intrínsecamente desordenadas: disección estructural de la oncoproteína E7 del virus del papiloma humano y de la fosfoproteína P del virus sincicial respiratorio”

RESUMEN

Las proteínas intrínsecamente desordenadas (IDPs) son proteínas funcionales que carecen de estructura secundaria y terciaria definidas en solución. Este tipo de categoría estructural se encuentra sobrerepresentada en proteínas virales, otorgándole multifuncionalidad y versatilidad en las interacciones que establecen. Debido a flexibilidad asociada a las distintas transiciones conformacionales que presentan, las IDPs son muy difíciles de caracterizar estructuralmente, debiéndose recurrir principalmente a técnicas de estudio en solución. En este trabajo se estudiaron dos modelos de IDPs virales: el dominio amino-terminal de la oncoproteína E7 (E7N) del virus del papiloma humano (HPV-16) y la fosfoproteína P del virus sincicial respiratorio humano (RSV).

En primer lugar, utilizando un abordaje de fragmentación en conjunto con medidas de dicroismo circular y resonancia magnética nuclear, se estudiaron las tendencias conformacionales del dominio E7N. Identificamos tres regiones con estructura secundaria local y transitoria: dos α-hélices dentro de las regiones conservadas CR1 y CR2, y una región que adopta hélice de poliprolinas tipo II (PII) dentro del CR2. Las dos α-hélices se encuentran pobladas en forma parcial, representando solo al 20% de la población de moléculas de E7N. Por otro lado, demostramos que la hélice-II presenta transiciones conformacionales entre PII y α-hélice moduladas por cambios de pH. Estas regiones de estructura local coinciden con los motivos lineales de interacción descriptos para este dominio, así como con sitios que experimentan modificaciones post-traduccionales. Estas características sugieren que dichas transiciones estructurales podrían modular la accesibilidad a estos motivos, regulando la interacción con sus proteínas blanco.

En segundo lugar, utilizando distintas proteínas recombinantes, demostramos por primera vez la naturaleza IDP de la proteína P de RSV. La misma contiene dos módulos intrínsecamente desordenados, los cuales presentan distinto contenido de desorden. Definimos al dominio C-terminal de P como un dominio con características en solución de tipo *pre-molten globule like IDP*, y al dominio N-terminal como un dominio con características de tipo *random-coil like IDP*. Describimos que P presenta un comportamiento térmico anómalo, conteniendo en el dominio C-terminal un elemento marginalmente estable modulable por variaciones de temperatura dentro del rango de temperaturas fisiológico.

En conclusión, estos ejemplos resaltan la flexibilidad funcional inherente de las IDPs, la cual les permite ser moduladas por cambios en el entorno. Esta capacidad permite a su vez explicar la plétora de procesos biológicos en los cuales se encuentran involucradas.

Palabras claves: Proteínas intrínsecamente desordenadas (IDPs), oncoproteína E7 HPV-16, dominio E7N, virus sincicial respiratorio humano, fosfoproteína P, flexibilidad conformacional, transiciones estructurales orden-desorden.

“Viral intrinsically disordered proteins: structural dissection of the human papillomavirus E7 oncoprotein and the respiratory syncytial virus P phosphoprotein”

ABSTRACT

The intrinsically disordered proteins (IDPs) are functional proteins that lack defined secondary or tertiary structure in solution. This type of structural category is overrepresented in viral proteins, giving multi-functionality and versatility in interactions. Because of the flexibility generated by the different structural transitions, the structural characterization of IDPs represents a challenging task, being necessary to use in solution techniques for their characterization. In this work we studied the structural and functional determinants of two IDP models: the human papilloma virus E7 aminoterminal IDP domain (E7N HPV-16) and the respiratory syncytial virus P phosphoprotein.

First, using a fragmentation approach in combination with circular dichroism and nuclear magnetic resonance experiments we studied the conformational tendencies within the E7N domain. We identified three regions with local secondary structure: two α-helixes within the CR1 and CR2 conserved regions, and a polyproline type II helix (PII) within the CR2. The two α-helixes were partially populated, representing only the 20% of the conformational ensemble within E7N. On the other hand, we demonstrated that the Helix-II presents a conformational transition between PII and α-helix triggered by pH changes. The small linear motifs as well as the post-translational modification sites are located within the mapped local secondary structural elements within E7N, suggesting that the structural transitions within these regions modulated the accessibility of the motives and in consequence regulates the interactions with target proteins.

In the second part of this thesis, we experimentally demonstrate for the first time the IDP nature of the P phosphoprotein, reveling the existence of two IDPs modules with different degree of disorder. The C-terminal domain behaves as a pre molten globule like IDP, whereas the N-terminal domain presents random coil-like IDP features. P presents a pronounced reversible secondary structure transition. We identified a marginal stable element within the C-terminal domain which could be tuned by temperature changes in the physiological range.

To conclude, these examples highlights the capability of the IDPs to be modulated by changes in their environment, bringing functional flexibility that may explain the plethora of biological processes where their are involved.

Keywords: Intrinsically disordered proteins (IDPs), HPV-16 E7 oncoprotein, E7N domain, Respiratory Syncytial Virus, P phosphoprotein, conformational ensemble, order-disorder structural transitions.

PRODUCCION CIENTIFICA

Los resultados presentados en esta tesis se encuentran publicados en los siguientes artículos:

-**NOVAL MG**, ESPERANTE S, MOLINA IG, CHEMES LB, de PRAT GAY G. Intrinsic Disorder to Order Transitions in the Scaffold Phosphoprotein P from the Respiratory Syncytial Virus RNA Polymerase Complex. (2016). *Biochemistry.* PMID: 26901160 (Published ahead of print).

-NOVAL MG, GALLO M, PERRONE S, SALVAY AG, CHEMES LB, de PRAT GAY G. Conformational dissection of a viral intrinsically disordered domain involved in cellular transformation. (2013). PLoS ONE. 8(9):e72760.

Durante el desarrollo de esta tesis, también se publicaron los siguientes artículos en colaboración con otros proyectos del laboratorio y un capítulo de libro:

-ESPERANTE SA, **NOVAL MG**, ALTIERI TA, de OLIVERA GA, SILVA JL, de PRAT GAY G. Fine modulation of the respiratory syncytial virus M2-1 protein quaternary structure by reversible zinc removal from its Cys(3)-His(1) motif. (2013). *Biochemistry.* 52(39):6779-89.

-CHEMES LB, **NOVAL MG**, SANCHEZ IE, de PRAT GAY G. Folding of a cyclin box: linking multitarget binding to marginal stability, oligomerization and aggregation of the retinoblastoma tumor suppressor AB pocket domain". *J. Biol. Chem*. (2013), 288, 26: 18923-18938.

-CHEMES LB, ALONSO LG, **NOVAL MG**, de PRAT GAY G. Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains. *Methods Mol Biol.* (2012), 895: 387-404. (capítulo de libro).

INDICE

Pag.

ABREVIATURAS 1

INTRODUCCIÓN 3

PARADIGMA DE “ESTRUCTURA-FUNCIÓN” DE PROTEINAS. 5

PROTEINAS INTRINSECAMENTE DESORDENADAS 6

Características Generales. 7

Clasificación de desorden intrínseco. 8

Caracterización experimental de desorden intrínseco. 10

Características de la estructura secundaria residual presente en IDPs. 12

Clasificación funcional de las IDPs. 13

Elementos de secuencia que median la interacción IDP-ligando. 13

Abundancia de IDPs en distintos sistemas biológicos. 15

DESORDEN INTRINSECO EN PROTEINAS VIRALES 16

Características generales del genoma viral. 16

IDPs y virus. 16

LA ONCOPROTEÍNA E7 DE LOS PAPILOMAVIRUS HUMANOS. 17

Los papilomavirus humanos. 17

HPVs: ciclo de vida, genoma y proteínas. 17

La oncoproteína E7 de HPV-16. 18

Propiedades bioquímicas y estructurales de HPV16 E7. 19

LA FOSFOPROTEÍNA P DEL VIRUS SINCICIAL RESPIRATORIO HUMANO. 22

El virus Sincicial Respiratorio Humano. 22

Clasificación filogenética. 22

Características generales de RSV: genoma y proteínas. 23

El complejo polimerasa de RSV 24

La fosfoproteína P. 25

Modificaciones post-traduccionales de P. 28

Interacciones con proteínas del complejo polimerasa. 28

RAZONAMIENTO, ALCANCE DEL TRABAJO Y OBJETIVOS. 31

RESULTADOS 33

CAPITULO I Disección estructural del dominio E7N de HPV-16. 35

DISECCIÓN ESTRUCTURAL DEL DOMINIO E7N DE HPV-16. 37

I.1.- CARACTERIZACIÓN DE LOS ELEMENTOS DE ESTRUCTURA PRESENTES EN E7N POR FRAGMENTACIÓN Y ESTABILIZACIÓN POR SOLVENTES. 37

I.1.1- Efecto del pH sobre el radio hidrodinámico del dominio E7N determinado por RMN. 38

I.1.2- Análisis de las propiedades hidrodinámicas de E7N por AUC. 40

I.1.3- Análisis del contenido de estructura secundaria local por dicroísmo circular. 43

I.1.4- Predicción de tendencia α-hélice dentro del dominio E7N. 44

I.1.5- Estabilización de estructura α-hélice por 2,2,2-trifluoroetanol (TFE). 45

I.1.6- Estabilización α-hélice por pH en el contexto de la proteína entera. 53

I.1.7- Estabilización de estructura secundaria en E7N por el detergente SDS. 54

I.1.8- Estructura de poliprolina tipo II en E7N. 56

I.2.- DETERMINACIÓN DE LOS ELEMENTOS DE ESTRUCTURA SECUNDARIA DE E7N POR RESONANCIA MAGNÉTICA NUCLEAR. 59

I.2.1- Asignación de las resonancias de RMN en E7N. 59

I.2.2- Identificación de α-hélices locales en E7N por análisis de CS. 61

I.2.3- Identificación de estructura PII en E7N por RMN. 63

DISCUSIÓN 67

Efecto del pH sobre la hélice-I, la hélice-II y la región PII. 67

Participación de las prolinas en la estabilización de elementos de estructura transitoria. 69

Los elementos de estructura definidos para E7N contienen a todos los motivos lineales de interacción descriptos para el dominio. 71

Modelo de las posibles transiciones conformacionales del dominio IDP E7N. 74

CAPITULO II Caracterización bioquímica y biofísica de la fosfoproteína P de RSV. 77

CARACTERIZACIÓN BIOQUÍMICA Y BIOFÍSICA DE LA FOSFOPROTEÍNA P DE RSV. 79

II.1.- ANÁLISIS HIDRODINÁMICO DE P Y LOS DISTINTOS FRAGMENTOS. 80

II.2.- LA FOSFOPROTEÍNA P DE RSV ES UNA IDP TIPO “MOLTEN GLOBULE”. 83

II.2.1- Análisis del contenido de estructura secundaria de P y de sus distintos módulos. 84

II.2.2- Análisis del comportamiento modular de P. 86

II.2.3- Análisis de cambios estructurales inducidos por temperatura. 87

Estudios al equilibrio. 87
Estudios cinéticos de la primera transición de desplegado. 91
II.2.4- Estudios de la estabilidad química de P. 92
Estudios al equilibrio. 92
Cinética del desplegado por Gdm.Cl. 95
II.2.5- Unión a la sonda fluorescente 8-anilino-1-naftalenosulfonato (ANS). 96
II.3. TRANSICIONES ESTRUCTURALES Y ESTUDIOS CONFORMACIONALES DE LOS FRAGMENTOS P_N Y P_C. 99
II.3.1- Estabilización estructural inducida por pH. 99
II.3.2- Estabilización de estructura α-hélice por TFE. 102
II.3.3- Integración de las tendencias conformacionales de los módulos de P. 103
DISCUSION 105
La fosfoproteína P de RSV es intrínsecamente desordenada. 105
Conservación dentro de la subfamilia *Pneumovirinae*. 106
P de RSV en el contexto de la subfamilia *Paramyxovirinae*. 109
Correlaciones funcionales de las transiciones estructurales estudiadas. 110

CONCLUSION GENERAL 113

CONCLUSIÓN GENERAL 115
(i) Plasticidad estuctural en IDPs. 115
(ii) Papel biológico de los elementos de estructura secundaria local. 116
(iii) Tipos de desorden intrínseco de proteínas. 117
(iv) IDP y Virus. 118

MATERIALES Y METODOS 121

APENDICE 145

REFERENCIAS 167

ABREVIATURAS

ADN Ácido desoxiribonucleico.

ANS 1-anilino-8-naftaleno-sulfonato

ARN Ácido ribonucleico.

AUC Ultracentrifugación analítica, por sus siglas en inglés.

CD Dicroísmo Circular, por sus siglas en inglés.

C-terminal Extremo carboxilo terminal

CR1 Región conservada 1

CR2 Región conservada 2

CS Desplazamiento químico, por sus siglas en inglés.

E7 Oncoproteína E7 de HPV

E7C Dominio C terminal de E7

E7N Dominio N terminal de E7

f/fmin Cociente de fricción

Gdm.Cl Cloruro de Guanidinio

HPV Virus del Papiloma Humano, por sus siglas en inglés.

HSQC Espectroscopia de coherencia cuántica única heteronuclear, por sus siglas en inglés.

IDP Proteína intrínsecamente desordenada, por sus siglas en inglés.

MBP Proteína de unión a maltosa, por sus siglas en ingles.

MoRF Elemento de reconocimiento molecular, por sus siglas en inglés.

MW Peso molecular, por sus siglas en inglés

N-terminal Extremo amino terminal.

ORF Marco abierto de lectura, por sus siglas en inglés.

pRb Supresor tumoral Retinoblastoma

PCR Reacción en cadena de la polimerasa, por sus siglas en inglés.

PII Hélice de poliprolinas de tipo II

PFG-NMR Resonancia magnética de gradiente de campo pulsado, por sus siglas en inglés.

RSV Virus sincicial respiratorio, por sus siglas en inglés.

RMN Resonancia magnética nuclear, por sus siglas en inglés.

R_S Radio hidrodinámico de Stokes

URR Región regulatoria no codificante, por sus siglas en inglés.

SDS Dodecilsulfato de sodio

SLiMs Motivo lineal de interacción, por sus siglas en inglés.

SLS Dispersión de luz estática, por su siglas en inglés.

TFE 2,2,2-trifluoroetanol

INTRODUCCION

PARADIGMA DE "ESTRUCTURA-FUNCIÓN" DE PROTEINAS.

En el año 1894, el premio Nobel alemán Hermann Emil Fischer descubrió que el α-metilglucosido y el β-metilglucosido, ambos obtenidos a partir de la metilación de la glucosa, exhibían especificidad por diferentes enzimas durante el proceso de hidrólisis. El isómero α era hidrolizado por la enzima "invertina" extraída de levaduras, mientras que el isómero β era hidrolizado por "emulsina" obtenida de almendras, no observándose hidrólisis cuando las enzimas eran intercambiadas. Estas observaciones llevaron a Emil Fischer a proponer el concepto de "llave-cerradura", donde una enzima y su sustrato debían poseer complementariedad geométrica (encajando como una llave a su cerradura), para que pueda ocurrir la catálisis enzimática [1]. Este concepto fue aceptado en el campo de la bioquímica de proteínas y por muchos años fue adoptado como paradigma, el cual propone que las proteínas deben poseer una única estructura tridimensional consolidada para ser funcionales. En 1930 Linus Pauling, entre otros, comenzó a investigar el proceso de desnaturalización de proteínas, observando que la pérdida de la estructura nativa de una proteína llevaba a la pérdida de su función, aportando evidencias a favor del paradigma de estructura-función [2]. Por otra parte, con el advenimiento de la cristalografía de rayos X para la resolución de estructuras proteicas y el consecuente aumento de la información estructural disponible, el paradigma de estructura-función se terminó de establecer en el campo [3, 4]. Las evidencias acumuladas arraigaron la noción de que las proteínas debían existir en un único estado nativo con una estructura terciaria y/o cuaternaria definidas para llevar a cabo su función biológica.

Sin embargo, a mediados del siglo XX comenzaron a surgir evidencias experimentales que contradecían el paradigma establecido. Por ejemplo, en 1950 McMeeking sugirió que "la caseína se encuentra en la leche en configuraciones desplegadas, siendo sensible a digestiones proteolíticas" [5]. Por otro lado, el glucagón monomérico, fue descripto en 1978 como "flexible y desplegado en solución acuosa, pero que presumiblemente podría plegarse al interactuar con su receptor" [6]. De todas formas, no fue hasta finales de los años 90 que empezó a ganar relevancia el concepto de que proteínas sin estructura definida en solución podrían poseer funcionalidad biológica. Muchas evidencias experimentales surgieron a partir del estudio de estados desplegados en solución acuosa. Por otra parte, la aplicación de la Resonancia Magnética Nuclear (RMN) al estudio de proteínas en solución permitió acceder a información estructural de alta resolución. Estos avances permitieron comenzar a comprender el papel de la flexibilidad conformacional y otorgarle una implicancia funcional a la misma. Varios ejemplos apoyando el concepto de la flexibilidad funcional fueron observados en proteínas principalmente asociadas a procesos de interacción proteína-proteína o proteína-ácido nucleico. En particular, algunos ejemplos son el dominio N-terminal de unión a Cdk2 del inhibidor de kinasas dependiente de ciclinas p21 [7], un fragmento de 130 residuos de la

proteína de unión a fibronectina de *Staphylococcus aureus* [8], y la región básica de unión a ADN del factor de transcripción de levaduras GCN4 [9, 10], entre otros [11].

Por otra parte, en 1998 se desarrolló un algoritmo de predicción de desorden, el cual permitió predecir desorden intrínseco a partir de la secuencia primaria de las proteínas [12]. En base a este predictor, se propuso que más de 15.000 proteínas en la base de datos Swiss Protein (Swiss-Prot) contenían regiones desordenadas de hasta 40 residuos consecutivos [12], sugiriendo que las proteínas con estructura tridimensional definida comprendían únicamente la mitad del proteoma. En su conjunto, las evidencias experimentales y computacionales mencionadas anteriormente, llevaron a desafiar el paradigma vigente de estructura-función, para incluir el nuevo concepto de desorden intrínseco en proteínas, y plantear el paradigma alternativo de “desorden-función” (Figura 1).

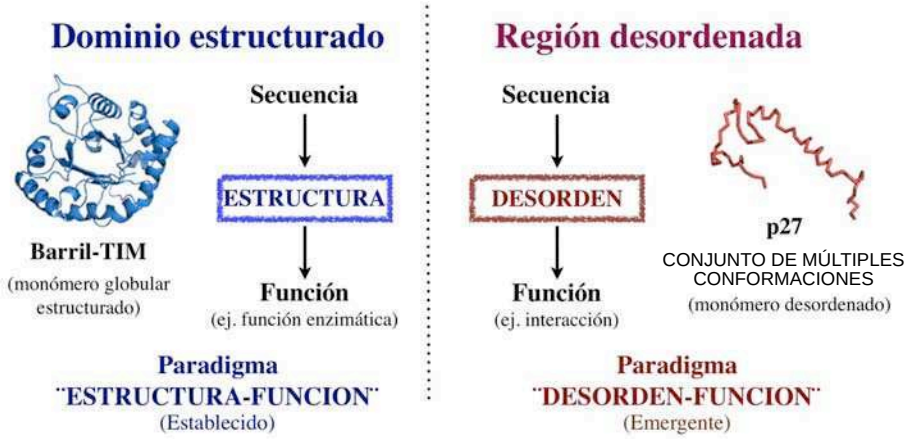


Figura 1. Representación esquemática de los paradigmas vigentes asociados a la versatilidad estructural de las proteínas. Adaptado de [14].

PROTEINAS INTRINSECAMENTE DESORDENADAS

Las proteínas intrínsecamente desordenadas (IDPs, por sus siglas en inglés) son proteínas funcionales a pesar de carecer de una única estructura secundaria y/o terciaria definida en solución. Debido a su naturaleza desplegada, frecuentemente las IDPs han sido identificadas erróneamente como estados desplegados tipo *random-coil*. Para éstos últimos se asume un “estado ideal” en el que cada aminoácido individual del polipéptido presenta grados de libertad para muestrear libremente todo el espacio estéricamente permitido del gráfico de *Ramachandram* [13]. En otras palabras, en este estado no existen diferencias en la energía libre de Gibbs de las distintas conformaciones que puede adoptar un polipéptido [13]. Las IDPs difieren de la conformación *random-coil* debido a que presentan regiones con tendencias estructurales locales tanto de estructura secundaria así como también, en algunos casos, cierto grado de estructura terciaria [14-16], lo cual restringe los posibles ángulos que puede adoptar el polipéptido. Las IDPs existen en la naturaleza como conjuntos dinámicos de múltiples

conformaciones que pueden experimentar diversas transiciones estructurales según el entorno físicoquímico en el que se encuentren [16, 17]. Funcionalmente, se las asocia a procesos regulatorios como transcripción, traducción y replicación, vías de señalización, control del ciclo celular, etc [16], formando parte de procesos que involucran interacción con distintos blancos (proteicos, ácidos nucleicos, membranas, etc), a diferencia de las enzimas que llevan a cabo funciones asociadas a reacciones químicas.

Previo a que sean consideradas como un único grupo de proteínas con características específicas, esta clase de proteínas dinámicas, flexibles y de gran importancia biológica fueron descriptas por múltiples grupos de investigación, acuñando diversos términos para definirlas, siendo algunos de ellos: proteínas flexibles [18], móviles [19], nativamente desnaturalizadas [20], nativamente desplegadas [21], nativamente desordenadas [22], proteínas camaleón [23], proteínas bailarinas [24], entre otros. Esta gran variedad de nombres generó la necesidad de unificar el término, dando origen al nombre actualmente aceptado de Proteínas Intrínsecamente Desordenadas o IDPs.

Características Generales.

Las IDPs, en comparación con las proteínas globulares, presentan una baja proporción de aminoácidos hidrofóbicos y/o aromáticos (Ej; Val, Leu, Ile, Met, Phe, Trp y Tyr), y una proporción alta de residuos polares y/o cargados (Ej; Ser, Gln, Pro, Asp, Glu, Arg, Lys, Gly y Ala) [25]. En un análisis realizado a partir de 275 proteínas estructuradas y 91 IDPs, se demostró que estas últimas presentan una relación carga neta/hidrofobicidad característica (alta carga neta y baja hidrofobicidad), la cual constituye una marca específica de este tipo de proteínas (Figura 2-A) [26]. La elevada carga neta en residuos cercanos en secuencia genera repulsiones electrostáticas, las cuales sumadas a la baja hidrofobicidad de sus secuencias, resultan en una reducción de la tendencia a adoptar estructuras compactas. Asimismo, la elevada proporción de residuos cargados hacen que estas proteínas presenten puntos isoeléctricos (*pI*) extremos, siendo principalmente proteínas ácidas o básicas (Figura 2-B) [26]. Por otra parte, estudios de desnaturalización térmica indican que las transiciones estructurales que experimentan las IDPs durante los cambios de temperatura son reversibles y no cooperativas [27].

Las IDPs poseen radios hidrodinámicos de Stokes (R_S) mayores a los esperados para proteínas globulares del mismo peso molecular, lo cual pone en evidencia su naturaleza extendida. Además, en muchos casos presentan movilidad electroforética anómala en geles de poliacrilamida en condiciones desnaturalizantes, observándose especies de peso molecular mayor al esperado. Este fenómeno se debe a que, por sus características bioquímicas, las IDPs interactúan con una menor cantidad de moléculas de detergente SDS. Por otra parte,

estas proteínas presentan susceptibilidad a degradación por proteasas, característica asociada a la ausencia de estructura compacta [16].

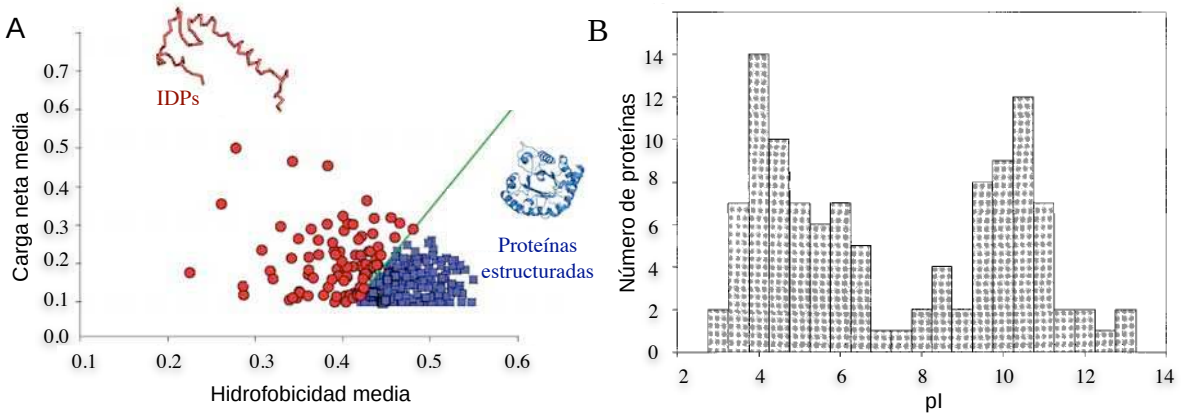


Figura 2. Características de las IDPs. A-Gráfico de carga *versus* hidrofobicidad comparativo para un conjunto de proteínas estructuradas (círculos azules) y un conjunto de proteínas IDPs (círculos rojos). La línea recta verde corresponde a la función <R>=2.743<H>-1.109, donde <R> y <H> representan la carga e hidrofobicidad media, respectivamente. B-Histograma representando la distribución de proteínas IDPs según sus puntos isoeléctricos (pI). Adaptado de [26].

Finalmente, las IDPs pueden poseer distinto grado de estructura, presentando elementos de estructura secundaria y/o terciaria transitorios, los cuales pueden ser estabilizados por interacciones con sus moléculas blanco, distintas condiciones de solvente, pH, temperatura o mediante modificaciones post-traduccionales [16].

Clasificación de desorden intrínseco.

Recientemente, diversas evidencias experimentales han demostrado de que las IDPs pueden presentar distinto grado de desorden intrínseco y por lo tanto no forman un grupo estructural homogéneo (Figura 3-A). Sin embargo, debido a la gradualidad del contenido de desorden intrínseco observado en la naturaleza, los límites entre clases de IDPs suelen ser difusos. Para esta tesis adoptamos la clasificación detallada a continuación. Nótese que algunos parámetros seleccionados son arbitrarios y tienen como fin facilitar la nomenclatura.

Las IDPs son clasificadas según el contenido residual de estructura secundaria y/o terciaria que poseen y el entorno fisicoquímico en el que se encuentren en las siguientes categorías: *IDPs tipo "molten globule"*, *IDPs tipo "pre molten globule"* e *IDPs tipo "random-coil"* [16].

A continuación se presentan las características principales de cada categoría.

- *IDPs tipo "molten globule".* Es un estado de desorden en el cual existe estructura secundaria nativa pero sin presentar la estructura compacta y cooperativa esperada para una proteína globular. Este estado posee características distintivas, como ser:

-R_S extendidos que corresponden a un aumento en el volumen de la proteína de un 50% respecto a su contraparte globular.

-Capacidad de unir a la sonda fluorescente 1-anilino-8-naftaleno-sulfonato (ANS).

-Accesibilidad a proteasas.

-Estructura secundaria estable.

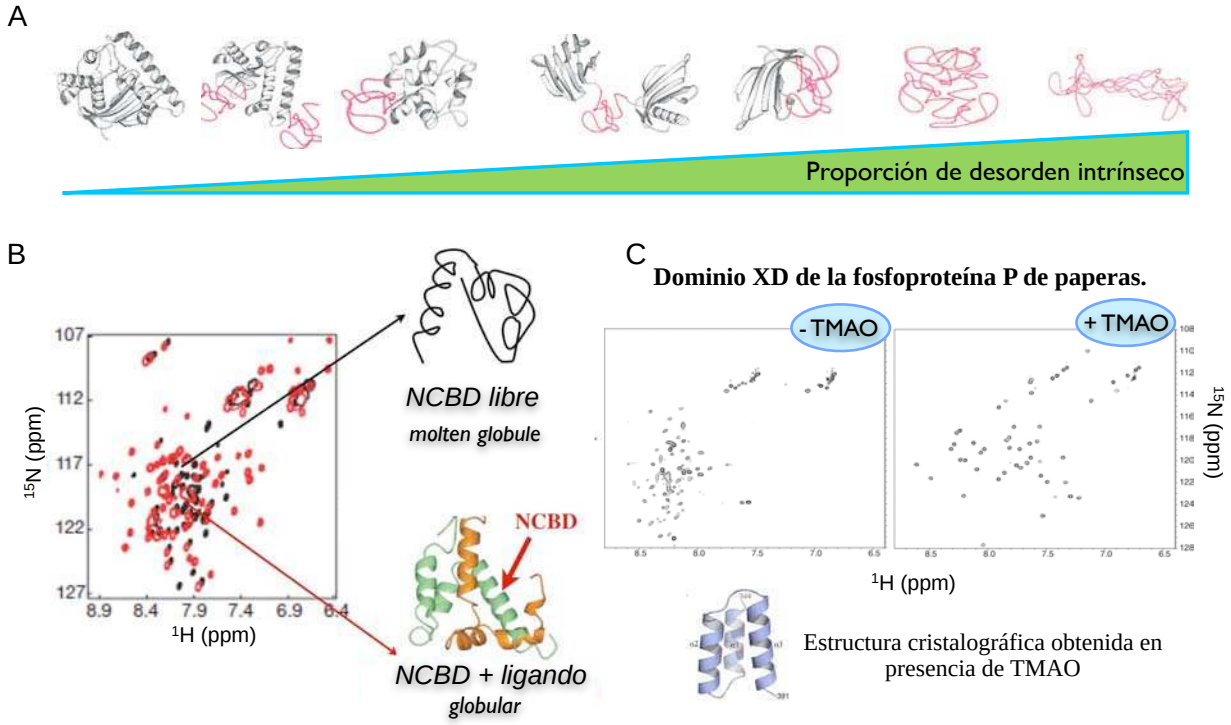


Figura 3. Caracterización de proteínas intrínsecamente desordenadas. A-Representación esquemática del gradiente de desorden intrínseco que puede existir en la estructura de una proteína. *B y C-Ejemplos de IDPs tipo Molten Globule. B*-Espectro de RMN bidimensional HSQC mostrando el ordenamiento estructural del dominio NCBD en presencia de su ligando (en rojo) en comparación a su estado libre (en negro) en donde adopta una conformación tipo *molten globule* [28]. C-Espectro de HSQC del dominio XD de la fosfoproteína P del virus de la parotiditis humana (paperas) en presencia y en ausencia de co-solvente *TMAO.* Se muestra la estructura cristalográfica del dominio XD obtenida en presencia de TMAO [29].

En la figura 3 -B y C se muestran dos ejemplos de este tipo de estado conformacional. Uno de ellos se corresponde con un estudio realizado sobre el dominio NCBD de la proteína de unión a CREB mediante RMN y experimentos de desplegado químico y térmico, demostrando que dicho dominio posee una conformación en solución de tipo *molten globule* en ausencia de ligando [28]. Sin embargo, al interactuar con el ligando específico ACTR, el dominio NCBD adopta una conformación globular (Figura 3-B, espectro rojo) [28]. El segundo ejemplo es un estudio realizado sobre el dominio XD de la fosfoproteína P del virus de la parotiditis humana (paperas), que media la interacción con la proteína de la nucleocápside. Se demostró por RMN, dicroísmo circular y calorimetría diferencial de barrido, que el dominio XD presenta características de tipo *molten globule* en solución y ausencia de ligando (Figura 3-C) [29], pero que adopta estructura compacta en presencia del cosolvente TMAO, condición en la que pudo ser cristalizado [29].

- *IDPs tipo "pre-molten globule"*. Es un estado extendido pero con estructura secundaria transitoria. Presenta un R_S en solución más extendido que el observado para *molten globules*, aunque puede interactuar con la sonda ANS, sugiriendo que a pesar de ser extendido existen regiones hidrofóbicas con estructura residual. Un ejemplo de proteínas tipo *pre-molten globule* es el inhibidor de quinasas dependiente de ciclinas (cdk) p21, donde el dominio de unión a cdk de p21 fue descripto como un dominio extendido pero con tendencia α-hélice residual en condiciones nativas [7].

- *IDPs tipo random coil.* Estado extendido con propiedades hidrodinámicas similares a las esperadas para una conformación *random-coil*. Posee muy bajo contenido de estructura secundaria, aunque puede sufrir transiciones conformacionales que lo lleva a estabilizar distintos tipos de estructura secundaria [16]. Este estado no une ANS. Un ejemplo de estas transiciones se observó para la proteína ACTR, la cual posee características extendidas y sin estructura definida en estado libre, pero adopta un plegamiento globular cuando forma complejo con una de sus proteínas blanco [28].

Caracterización experimental de desorden intrínseco.

Las proteínas IDPs pueden ser estudiadas mediante una variedad de técnicas biofísicas y abordajes experimentales. Debido a su naturaleza extendida y dinámica las técnicas utilizadas se basan principalmente en estudios de sus propiedades estructurales en solución. Algunas de las técnicas más utilizadas para su caracterización *in vitro* son:

- Análisis de la estructura secundaria por dicroísmo circular al UV lejano (CD) [30]. Esta técnica se basa en la absorción diferencial de luz circularmente polarizada hacia la derecha y hacia la izquierda a una determinada longitud de onda. Para que un cromóforo presente bandas de CD, este debe encontrarse en un entorno asimétrico. En el caso de las proteínas los cromóforos más comunes son los grupos amida del enlace peptídico y los residuos aromáticos, los cuales presentan una geometría simétrica debido a su naturaleza planar. El fenómeno de CD se genera cuando estos cromóforos se encuentran en un ambiente asimétrico generado por los distintos tipos de plegamiento proteico. Las proteínas presentan bandas de CD en dos regiones espectrales diferentes: la región UV lejano (175-260 nm) y la región UV-cercano (260-300 nm). En particular, el análisis de las IDPs se realiza principalmente analizando las bandas de CD de las regiones del UV-lejano. Esta región da información sobre el contenido de estructura secundaria de una proteína, principalmente basada en la conformación asimétrica generada por la cadena carbonada del polipéptido. En general, los espectros de CD de proteínas resultan de una combinación de las bandas espectrales de los distintos tipos de estructura secundaria que la componen, dando lugar a espectros complejos con máximos y mínimos en posiciones diferentes a las esperadas para proteínas con elementos de estructuras secundaria puras (ver próxima sección). En

particular, los espectros de CD típicos de IDPs carecen de componentes de estructura secundaria canónica, presentan un mínimo definido alrededor de 200 nm, y se caracterizan por la ausencia de mínimos en la región cercana a la banda a 220 nm (Figura 4). Sin embargo, los espectros de CD de las IDPs pueden presentar aportes de distintos tipos elementos de estructura secundaria (ver próxima sección).

Esta técnica se ha convertido en un método ampliamente utilizado para el análisis de IDPs principalmente por su sensibilidad para detectar diferentes transiciones conformacionales involucrando cambios en estructura secundaria [30].

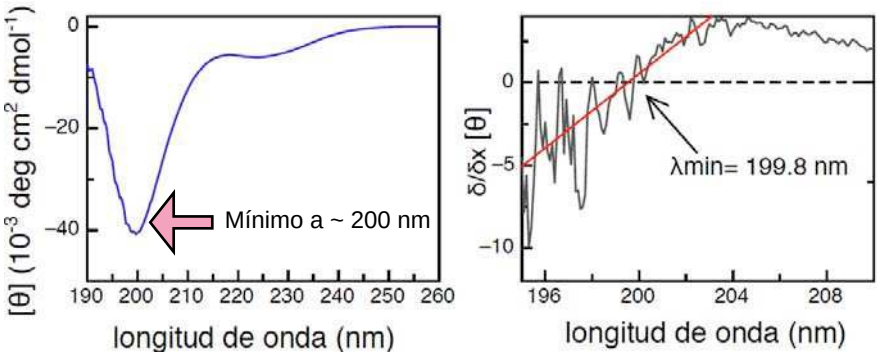


Figura 4. Espectro de CD característico de una proteína IDP. Determinación de la posición del mínimo a partir del cálculo de la primer derivada del espectro de CD (panel derecho). La posición del mínimo del espectro se corresponde con el punto en donde la derivada es igual a cero.

- Análisis de propiedades hidrodinámicas por cromatografía de exclusión molecular *(SEC,* por sus siglas en inglés) [31], dispersión de luz estática *(SLS,* por sus siglas en inglés) [32], ultra-centrifugación analítica *(AUC,* por sus siglas en inglés) [33], RMN de gradiente de campo pulsado *(PFG-RMN)* [34], entre otros.

- Dentro de los estudios de resolución atómica, la principal técnica de estudio para IDPs es la Resonancia Magnética Nuclear (RMN). El RMN, como ya hemos mencionado anteriormente, representa una de las técnicas más utilizadas para obtener información cuantitativa y a nivel residual de las proteínas IDPs en solución. Una caracterización general de una IDP puede ser inferida a partir del análisis de espectros monodimensionales (1D) o a partir de experimentos bidimensionales (2D). Estos últimos son los más frecuentemente utilizados para el estudio de IDPs en solución, ya que brindan el punto de partida para la asignación de las resonancias de cada residuo, siendo fundamentales para la obtención de información estructural a nivel residual [35]. Actualmente, a partir de estudios de RMN dentro de la célula, se está comenzando a obtener información estructural de las IDPs en condiciones endógenas [36]. Los resultados estructurales obtenidos por RMN pueden ser complementados con una combinación de la técnica de dispersión de rayos X de ángulo chico (SAXS, por sus siglas en inglés) y medidas de dinámica molecular, para obtener información completa del conjunto de conformaciones que puede poblar una IDP [37].

- Mapeo de regiones flexibles mediante el estudio de la accesibilidad a digestión por proteasas seguido por geles y por espectroscopía de masas *(MALDI-TOF)* [16].
- Caracterización de la estructura terciaria y sus transiciones estructurales, por espectroscopía de fluorescencia (incluyendo unión a sondas fluorescentes, FRET, espectroscopía de correlación de fluorescencia, experimentos de molécula única, etc) [38].

Características de la estructura secundaria residual presente en IDPs.

Como se mencionó anteriormente, las IDPs presentan distintos grados de estructura secundaria local. A lo largo de esta tesis se analizarán diversas transiciones conformacionales de distintos tipos de estructura secundaria, por lo que a continuación se describirán brevemente las características más relevantes de cada tipo de estructura.

α-hélice. Representa la conformación más predominante adoptada por un polipéptido. En general para una *α*-hélice típica cada giro está conformado por 3.6 residuos presentando una extensión de 5.4 Å por giro. Los ángulos dihedros de la cadena carbonada Φ y Ψ presentan valores de ~ -60º y -50º, respectivamente (Figura 5-A). Una característica importante de este tipo de estructura es que el grupo amida de un aminoácido en la posición *i* forma uniones tipo puente de hidrógeno con el grupo carbonilo del aminoácido en posición *i+4*. El espectro típico de dicroísmo circular para una estructura α-hélice canónica presenta dos mínimos característicos, uno a 208 nm y otro a 220 nm, respectivamente y una banda de valores de elipticidad positiva a 195 nm (Figura 5-B, línea verde).

Lámina-β. Consiste en dos o más cadenas-*β* unidas entre sí a través de interacciones tipo puente de hidrógeno. Esta estructura consiste en una cadena carbonada extendida cuyos ángulos dihedros Φ y Ψ pueden adoptar las posiciones -119º y 113º (lámina-*β* paralela) o las posiciones -139º y 135º (lámina-*β* antiparalela) (Figura 5-A). El espectro de CD típico de una lámina-*β* presenta un mínimo alrededor de 218 nm y un máximo alrededor de 195 nm (Figura 5-B, línea azul).

Hélice de poliprolinas tipo II (PII). Es una conformación predominante en muchas IDPs así como también en polipeptidos desplegados en agentes caotrópicos como UREA o Cloruro de guanidinio (Gdm.Cl) [39]. Se caracteriza por ser una hélice levógira y extendida, la cual presenta 3 residuos por giro [40]. La extensión ocupada por giro es de 9.3 Å, mayor a los 5.4 Å estimados para una α-hélice típica. Los ángulos de torsión Φ y Ψ adoptan las posiciones -75º y 145º, respectivamente, adoptando una simetría rotacional con forma de prisma triangular [40]. Debido a su carácter extendido la estructura PII no posee patrones de interacción tipo puente de hidrógeno en contraste con lo observado para las α-hélices. La estructura PII se encuentra en proteínas con regiones ricas en el prolinas, así como también en regiones con un elevado contenido de residuos cargados (Asp, Glu, Lys, Arg) ya que, para éstos últimos, las repulsiones

electrostáticas de las cadenas laterales cargadas favorecen la formación de este tipo de hélices extendidas [40]. El espectro de CD típico de PII se caracteriza por presentar una banda de elipticidad positiva alrededor de 218 nm y un mínimo a 198 nm (Figura 5-B, línea roja) [30]. Este tipo de conformación es estabilizada a bajas temperaturas llevando a un aumento tanto de la banda negativa a 198 nm, así como también de la banda positiva a 218 nm. El análisis del espectro diferencia entre dos temperaturas extremas (5 y 85 ºC) facilita la visualización de la estabilización de PII (Figura 5-B, inserto).

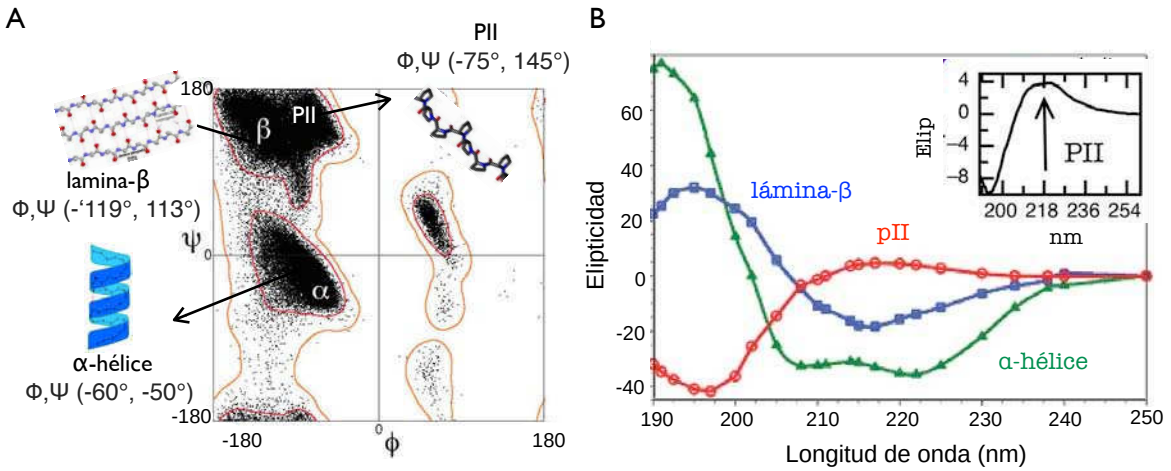


Figura 5- Elementos de estructura secundaria. A-Gráfico de *Ramachandram* indicando la posición de los ángulos dihédros Φ y Ψ característicos de los distintos tipos de estructura secundaria. B-Espectros de CD al UV lejano canónicos. En el inserto se muestra el espectro diferencia entre los espectros de CD a 5 y 85 °C de la proteína PII. La flecha indica el máximo a 218 nm característico de estructura PII.

Clasificación funcional de las IDPs.

Las proteínas o dominios intrínsecamente desordenados pueden cumplir funciones como “cadenas entrópicas” o como elementos de “reconocimiento molecular”. Las regiones que cumplen función de cadenas entrópicas no están asociadas a interacciones, sino que su función radica principalmente en otorgar flexibilidad y plasticidad sobre dicha región de la proteína gracias a su contenido de desorden intrínseco, por ejemplo en regiones espaciadoras de dominios [41]. Por otra parte, la función de reconocimiento molecular esta asociada a todo proceso de interacción IDP-ligando, donde los ligandos pueden ser: proteínas, ácidos nucleicos, moléculas pequeñas, iones, etc.

Elementos de secuencia que median la interacción IDP-ligando.

Actualmente, existen dos conceptos similares que hacen referencia a las regiones de las IDPs que median la interacción con su ligando. Estos son: los “motivos lineales de interacción” (SLiM) y los “elementos de reconocimiento molecular” (MoRFs). Ambos conceptos

son esencialmente análogos, y se ha visto para muchos ejemplos de pares de interacción IDP-ligando que los MoREs y los SLiM se superponen [42].

*Motivos lineales de interacción (*SLiMs*).* Los SLiMs son segmentos cortos de entre 3-20 residuos ubicados dentro de regiones intrínsecamente desordenadas que median las interacciones entre las IDPs y sus ligandos [43]. Estos motivos están involucrados en establecer contactos con las proteínas blanco y están principalmente enriquecidos en residuos cargados (Asp y Arg), Pro y/o residuos hidrofóbicos (Trp, Leu, Cys y Tyr) y depletados de Gly y Ala [43].

Elementos de reconocimiento molecular (MoRFs). Los MoRFs son segmentos de 10 a 70 residuos que intervienen en el reconocimiento de un ligando [44]. Están asociados principalmente a un elemento estructural que experimenta transiciones del tipo desorden-orden durante el proceso de interacción [44]. Existen tres tipos de MoRFs: α-MoRFs, β-MoRFs e ι-MoRFs, los cuales forman α-hélices, láminas-β o estructura irregular en el complejo, respectivamente.

A continuación se describen dos ejemplos que resaltan la plasticidad estructural de las regiones que median interacción entre las IDPs. Se observa que una misma región de secuencia de la proteína p53 puede interactuar con al menos 4 blancos celulares distintos. Cuando se analizó por cristalografía la conformación que adopta esta región en complejo con cada uno de los cuatro blancos proteicos, se observó que según el blanco al que está asociado, el motivo puede adoptar distintas estructuras (hélice extendida, α-hélice, lámina-β) resaltando la plasticidad conformacional de las IDPs (Figura 6-A) [45]. Un comportamiento parecido se observó para el dominio C terminal de transactivación IDP del factor inducible por hipoxia (HIF1-α). En condiciones de normoxia dicho factor es hidroxilado en la Ans803 por la enzima hidroxilasa FIH formando un complejo con FIH donde los residuos 795-805 adoptan una conformación extendida (Figura 6-B). Sin embargo, en condiciones de hipoxia la Asn803 no es hidroxilada y HIF 1-α interactúa con la proteína TAZ1, donde se observa que la región 795-805 adopta estructura α-hélice en complejo (Figura 6-B) [46].

"*Complejos Fuzzy":* interacción especifica sin plegamiento completo. Existen numerosos ejemplos que muestran que las proteínas IDPs pueden interactuar con los distintos blancos celulares a través del reconocimiento molecular [45] o tras experimentar transiciones de desorden orden luego de la unión con su ligando [44]. Sin embargo, se han acumulado ejemplos en donde se ha observado que la heterogeneidad conformacional de las regiones desordenadas se mantiene al formar del complejo con un ligando [47], y cuya flexibilidad es requerida para su función, éstos complejos son conocidos como Complejos *Fuzzy* [48].

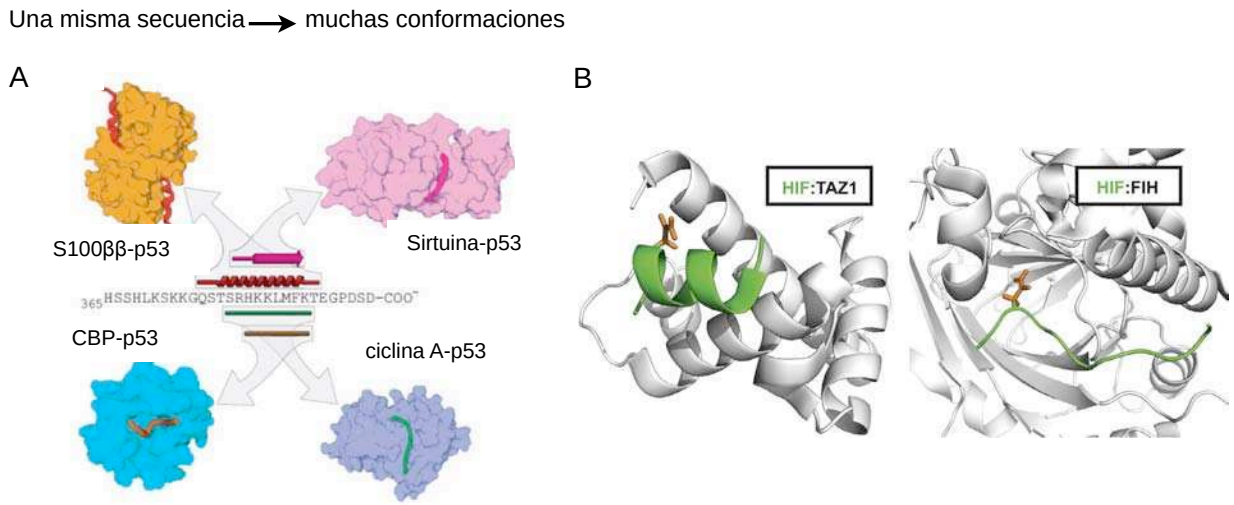


Figura 6. Ejemplos de la plasticidad estructural de regiones IDPs. A-Esquema de la región C terminal de p53 IDP indicando que una misma región puede adoptar múltiples conformaciones dependiendo del blanco al que este unido. B-En verde se muestran los residuos (795-805) del dominio C terminal IDP de HIF-1α (verde), adoptando conformación α-hélice al formar complejo con TAZ1 o adoptando conformación extendida al formar complejo con FIH.

Abundancia de IDPs en distintos sistemas biológicos.

Múltiples estudios bioinformáticos demostraron que el desorden intrínseco en proteínas se encuentra representado en los tres dominios de la vida: bacteria, archea y eucaria. En bacterias y archeas se ha descripto que el proteoma presenta entre un 18 y un 30% de proteínas o regiones intrínsecamente desordenadas mayores a 30 residuos, mientras que los proteomas eucariotas poseen entre un 35 y un 50% [49]. Varios trabajos independientes observaron una correlación entre la complejidad del organismo y la frecuencia y el largo de las regiones desordenadas de sus proteomas [50] [51], sugiriendo que el desorden intrínseco favorece la diversificación de funciones de una misma proteína.

Por otra parte, el contenido de desorden intrínseco en proteomas virales representa el 30% de los mismos [52]. Sin embargo, se ha observado que dicho porcentaje varía sustancialmente entre distintas familias de virus (entre el 3 y el 30 %) [53]. En particular, las dos familias que contienen a los sistemas de estudio de esta tesis, familia *Papillomaviridae* y familia *Paramyxoviridae,* presentan aproximadamente entre un 20% y un 12% de desorden intrínseco, respectivamente [53].

DESORDEN INTRINSECO EN PROTEINAS VIRALES

Características generales del genoma viral.

Los virus poseen genomas pequeños y compactos, los cuales en general codifican para no más de 15 a 20 productos proteicos. Debido a las características de su maquinaria transcripcional algunos genomas virales, principalmente los de los virus con genoma a ARN, son muy sensibles a acumular mutaciones con tasas estimadas de cambio de base por generación de 10^{-3} a 10^{-5} [54], en comparación con las tasas estimadas para la célula huésped (~10^{-8}) [54]. En algunos virus, se ha descripto la existencia de marcos abiertos de lectura (ORFs, por sus siglas en ingles) solapados, lo cual implica que una simple mutación podría afectar a más de una proteína. Los virus poseen la capacidad de infectar al huésped, secuestrar la maquinaria celular, replicar en distintos tipos de células y organismos, y todo con genomas pequeños que codifican para pocos productos proteicos. Frente a esto, surge el siguiente interrogante: *¿Como pueden los virus, los cuales codifican para tan poca cantidad de proteínas, infectar y desregular la maquinaria celular tan eficientemente?*

IDPs y virus.

La elevada flexibilidad estructural y funcional conferida por la presencia de IDPs en los proteomas virales permite dar una respuesta parcial a este interrogante. El desorden intrínseco les confiere a estas proteínas una elevada flexibilidad conformacional, la cual se traduce en que una misma proteína puede adoptar múltiples conformaciones según el entorno en el que se encuentre, por ejemplo, unida a un blanco específico, en una determinada localización subcelular, o en el contexto de modificaciones post-traduccionales, etc.

Durante el ciclo de vida de los virus, las proteínas virales interactúan con múltiples blancos celulares “secuestrando” la maquinaria celular para poder efectuar su ciclo de replicación. Este proceso es favorecido por la flexibilidad aportada por el desorden intrínseco, así como también por la presencia de SLiMs que median interacciones y se encuentran principalmente en regiones desordenadas. Diversos casos estudiados sugieren que hay un alto contenido de motivos lineales en proteínas virales [55]. Se propone que los virus secuestran la maquinaria celular a través de “imitar” motivos lineales de interacción típicos eucariotas compitiendo con las proteínas celulares por los distintos blancos [55]. Estos antecedentes sugieren que el desorden intrínseco en proteínas virales representa una ventaja adaptativa frente a la necesidad de utilizar eficientemente material genético limitado, de forma análoga al *splicing* alternativo o al solapamiento de genes en virus.

LA ONCOPROTEÍNA E7 DE LOS PAPILOMAVIRUS HUMANOS.

Los papilomavirus humanos.

Los papilomavirus humanos (HPV, por sus siglas en inglés) son virus sin envoltura de membrana con genoma de ADN doble cadena, circular y de tamaño pequeño (aproximadamente 8 Kpb). Actualmente se conocen más de 150 tipos virales distintos, a los cuales se los clasifica filogenéticamente en 5 géneros: alfa, beta, gamma, nu y mu. El género alfa contiene los tipos de HPVs de mayor relevancia médica, incluidos los asociados a cáncer cervical. Los HPVs poseen tropismo epitelial e infectan principalmente epitelios estratificados los cuales, según las prognosis clínicas de las lesiones a las que se asocian, pueden clasificarse como de *alto riesgo* o de *bajo riesgo*. Los HPVs de *bajo riesgo* (HPV-11 y HPV-6, entre otros) están asociados a lesiones benignas mientras que los de *alto riesgo* (HPV-16, HPV-18, HPV-45, entre otros) se encuentran asociados al desarrollo de lesiones pre-cancerosas que pueden progresar hacia cáncer cervical. El cáncer cervical o cáncer de cuello de útero constituye el 25% de los cánceres femeninos y es la segunda causa de muerte por cáncer en mujeres a nivel mundial [56]. En Argentina en particular, la tendencia indica un aumento en la incidencia de cáncer cervical en los últimos años registrando 5000 casos nuevos en la población femenina en 2012 [57]. Esta observación implica que a pesar de la existencia de las vacunas profilácticas contra HPV incluidas en el plan de vacunación nacional, es importante continuar con estudios que permitan comprender los mecanismos de acción del virus tanto a nivel estructural como molecular con el fin de diseñar nuevas drogas y vacunas terapéuticas.

HPVs: ciclo de vida, genoma y proteínas.

La infección por HPV se produce cuando las células de la capa basal de un epitelio plano estratificado entran en contacto con una partícula viral, principalmente debido a pequeñas micro lesiones del epitelio (Figura 7-A).

El ciclo de vida del virus se encuentra ligado al estadio de diferenciación del epitelio estratificado. Los primeros genes virales que se transcriben son los llamados tempranos (E, por sus siglas en inglés), cuya expresión se registra en células del estrato basal hasta el estrato córneo (Figura 7-A). En los estratos intermedios se produce la replicación del ADN. Por otra parte, a medida que avanza la diferenciación del epitelio, comienzan a expresarse las proteínas de expresión tardía L1 y L2 (L, por sus siglas en inglés), las cuales conforman la cápside viral (Figura 7-A). Finalmente, en el estrato granuloso maduran los viriones, que se desprenden al descamarse los queratinocitos terminales del estrato córneo (Figura 7-A), propagando la infección viral.

El genoma circular de los HPVs puede dividirse en tres regiones: una región regulatoria no codificante (URR, por sus siglas en inglés) la cual contiene los elementos regulatorios de la replicación y de la transcripción del genoma; y dos regiones que codifican para los genes de expresión temprana (E1, E2, E4, E5, E6 y E7) y para los genes de expresión tardía (L1 y L2) (Figura 7-B).

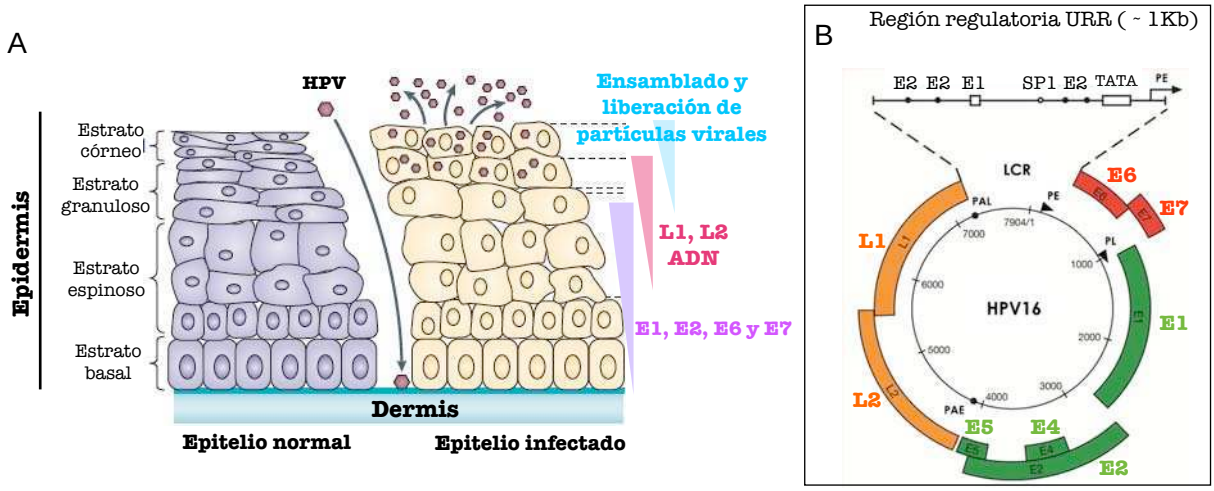


Figura 7. Esquema general del ciclo de vida y organización genómica de los HPVs. A-En la figura se detallan las características típicas del epitelio estratificado normal *versus* infectado. En forma de gradiente se representa la expresión diferencial de las distintas proteínas a lo largo de las distintas capas del epitelio. B-Genoma circular de HPV-16 en donde se detallan las posiciones de la región regulatoria URR y los distintos ORFs. En rojo y verde están indicados los ORFs de las proteínas de expresión temprana, mientras que en naranja se indican los ORFs de las proteínas de expresión tardía.

Los HPVs no codifican para las enzimas y factores proteícos necesarios para la transcripción y replicación de su genoma, lo cual implica que deben utilizar la maquinaria de replicación de la célula huésped para llevar a cabo su ciclo de vida. Dado que los HPVs poseen tropismo por células diferenciadas, las cuales no se encuentran en replicación activa, estos virus poseen dos proteínas de expresión temprana llamadas E6 y E7 que fuerzan la entrada en fase S del ciclo celular [58]. Estas proteínas son oncogénicas, ya que la expresión de las mismas induce inmortalización y transformación celular en distintos tipos celulares de ratón y humanos [58].

La oncoproteína E7 de HPV-16.

E7 de HPV-16 es una proteína pequeña de 98 aminoácidos, cuya localización puede ser tanto nuclear como citosólica [59]. Esta proteína cumple un papel central en la transformación celular, ya que se ha demostrado en estudios *in vitro* que su expresión lleva a la inmortalización de queratinocitos [60]. Su principal función está asociada a la interacción con el supresor de tumores retinoblastoma (pRb), desestabilizando el complejo pRb-E2F y produciendo la liberación del factor general de la transcripción E2F que regula genes de fase S que promueven la progresión del ciclo celular [61].

Propiedades bioquímicas y estructurales de HPV16 E7.

E7 fue descripta por primera vez como IDP por nuestro grupo en el año 2002, donde se demostró que presenta un radio hidrodinámico extendido y una elevada plasticidad conformacional que se traduce en la capacidad de adoptar diversas estructuras en presencia de distintas condiciones de pH y de solvente [62]. Posteriormente, se propuso que el dominio que le confería este carácter IDP a la proteína E7 era el dominio N terminal [63], siendo confirmada esta observación experimentalmente por nuestro laboratorio [64].

La proteína E7 esta compuesta por dos dominios conservados [65]: un dominio N terminal intrínsecamente desordenado (E7N) [63] y un domino C terminal globular (E7C), ambos conectados mediante una región bisagra rica en residuos de prolinas (Figura 8.A) [66]. El dominio E7N puede subdividirse en dos regiones conservadas CR1 (residuos 1-15) y CR2 (residuos 16-40), las cuales median la interacción con más de cuarenta blancos celulares y virales (Tabla I). Las regiones CR1 y CR2 comparten identidad de secuencia con regiones análogas de la proteína viral E1A de Adenovirus, así como también con la proteína del virus SV40 *Large T antigen* (Figura 8.A).

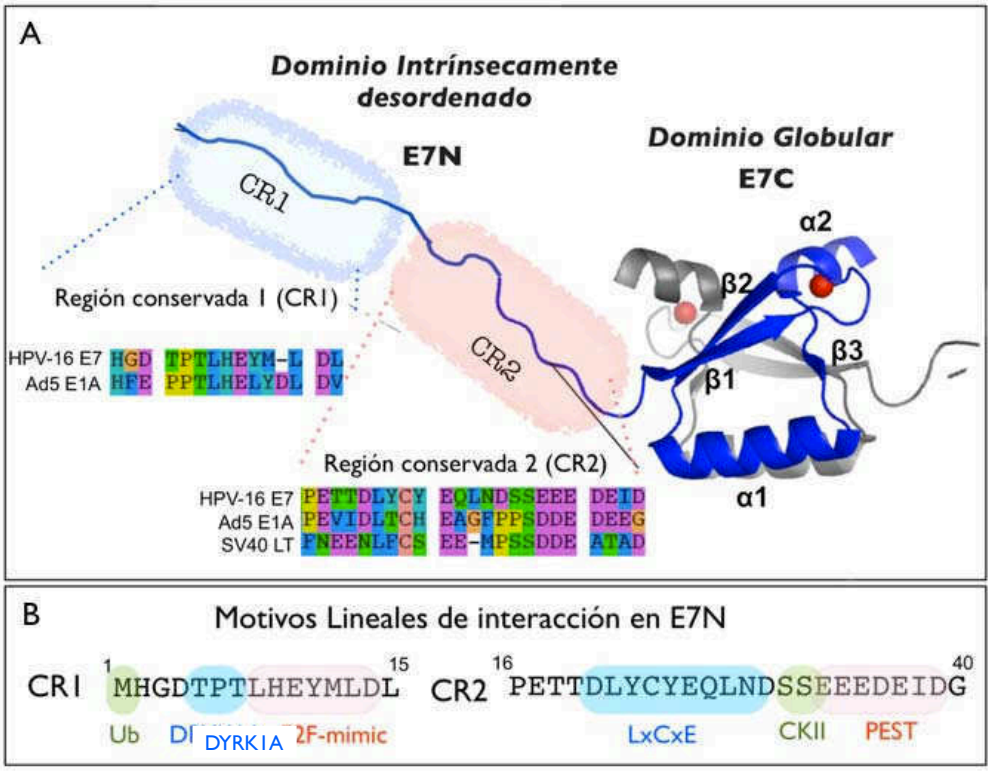


Figura 8. Representación esquemática de la oncoproteína E7 de HPV. A- Estructura por RMN del dominio E7C de HPV-45 (PDB ID: 2F8B). Cada monómero esta representado en azul y en gris, respectivamente y en rojo se indican los átomos de Zn^{2+} estructurales. El dominio E7N IDP de un monómero se muestra como cinta, se detallan las posiciones de las regiones conservadas (CR1 y CR2) y las secuencias de CR1 y CR2 de HPV-16 en comparación con la de otros virus a ADN tumorigénicos. B-Detalle de los motivos lineales de interacción identificados en las regiones CR1 y CR2 de HPV-16.

Se han identificado 5 motivos lineales de interacción contenidos en el dominio E7N [65] (Figura 8.B). Debido a que este dominio de 40 residuos se encuentra densamente poblado por dichos motivos (Figura 8-B), se deduce que los mismos podrían estar involucrados en la interacción con los blancos descriptos para el mismo (Tabla I).

Dentro de la región conservada CR1 se localiza un motivo de fosforilación para la quinasa DYRK1A y el motivo de interacción “E2F-mimic” que interfiere con la interacción entre pRb y E2F; además de un motivo de ubiquitinación [65] (Figura 8-B). Por otra parte, la región CR2 contiene al motivo “LxCxE” que media la interacción con el dominio AB de pRb de mayor afinidad; dos sitios de fosforilación para la enzima caseina kinasa II (CKII) y una región ácida que contiene el motivo de degradación PEST que regularía la interacción con Rb, así como también la degradación de E7 [65].

Tabla I. Blancos celulares que interactúan con el dominio E7N.

Proteínas blanco	Función asociada	Región de interacción en E7N	Cita
pRb, p130	Desestabilización e interrupción del complejo con E2F	CR1 y CR2	[165], [153]
P107	Disrupción del complejo con E2F	CR2	[153]
UBCH7, Skp2	Ubiquitinación de E7	CR1	[154]
p600	Transformación celular	CR1	[155]
pCAF	Disminución de la actividad acetil transferasa *in vitro*	CR1	[156]
FHL2	Transformación celular	CR1 y CR2	[157]
IRF-1	Inhibición de la respuesta INF	CR1 y CR2	[158]
nm23-H1, 2	Desconocida	CR1 y CR2	[159]
F-actina	Desconocida	CR2	[160]
CDKI, CDK2, Ciclina E2, Ciclina A2, Ciclina E1	Transformación celular	CR2	[161]
CKII	Fosforilación de E7N en S_{31} y S_{32}	CR2	[162]
E2F4, IRF-9	Transformación celular	CR2	[152], [151]
RAN	Desconocida	CR2	[163]
Smad2	Desconocida	CR2	[164]

Por otra parte, el dominio globular E7C presenta una topología β1β2α1β3α2 (Figura 8-A), es el dominio responsable de la dimerización de E7 y posee dos motivos conservados CxxC que coordinan Zn^{2+}.

E7 forma principalmente homodímeros débiles en solución (Kd aproximada de 1 µM), aunque también se observó que presenta múltiples equilibrios conformacionales, incluyendo equilibrios con dímeros globulares, monómeros y tetrámero oxidado (Figura 9). En particular, E7 puede formar oligómeros esféricos estables (E7SOs) cuando se quela el Zn^{2+} coordinado [67]. Los E7SOs presentan actividad chaperona *in vitro* [68] y poseen el dominio intrínsecamente desordenado E7N expuesto al solvente, lo cual aporta solubilidad al oligómero [67]. Por otra parte, se ha demostrado la existencia de los E7SOs *in vivo*, siendo identificados en el citosol de células transformadas, así como también en tejido canceroso de pacientes [69], donde se postuló que sería esta especie la que podría interactuar con múltiples blancos (Figura 9).

Contrario a lo esperado para las proteínas intrínsecamente desordenadas, los dominios E7N y E7C presentan un elevado grado de conservación de secuencia, y la misma proporción de residuos que co-evolucionan [69]. Esto sugiere que los elementos de secuencia de E7N son funcionalmente importantes para la proteína y por lo tanto, muy conservados.

Un trabajo reciente demostró una evidencia de acoplamiento funcional entre dos residuos de E7 (Cys 24 y Cys 59) pertenecientes a los dominios E7N y E7C, respectivamente. Se observó que la presencia de la Cys 59 es esencial para la funcionalidad de la proteína E7 ya que este residuo interviene directamente en la regulación del estado de oxidación de la Cys 24 perteneciente al motivo de unión a pRb LxCxE [70]. Otro ejemplo de acoplamiento funcional fue observado dentro de E7N a partir de experimentos de co-evolución, donde se definió que los motivos LxCxE y la región ácida CKII-PEST se encuentran cerca en secuencia, y presentan señales de co-evolución [65].

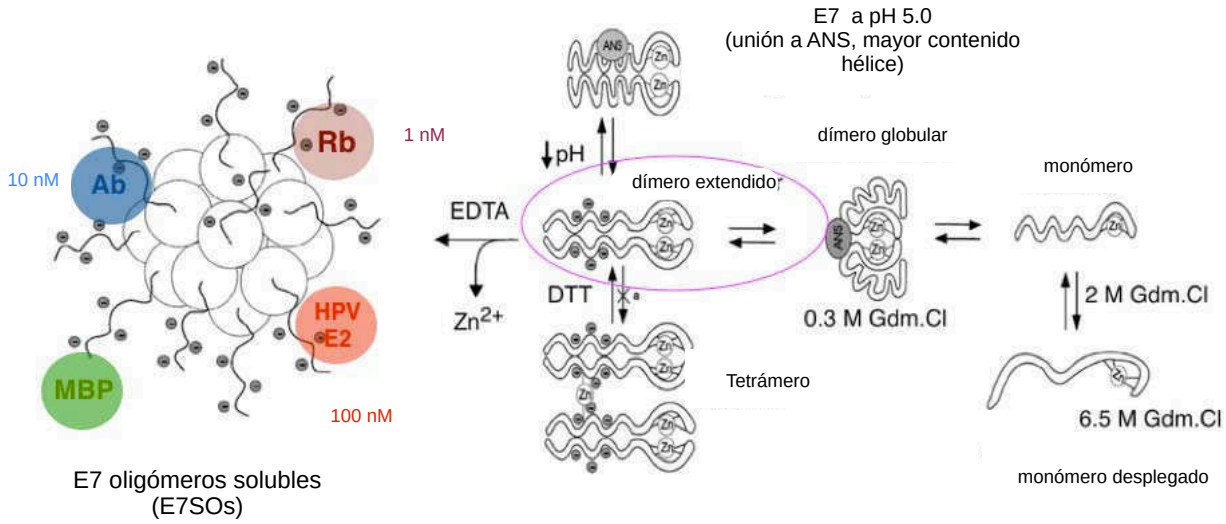


Figura 9. Equilibrios conformacionales descriptos para E7 y ejemplos de las interacciones estudiadas en el laboratorio.

LA FOSFOPROTEÍNA P DEL VIRUS SINCICIAL RESPIRATORIO HUMANO.

El virus Sincicial Respiratorio Humano.

El virus Sincicial Respiratorio Humano (RSV) es el principal agente causal de infecciones del tracto respiratorio inferior (bronquiolitis y neumonia) en niños de todo el mundo [71]. Según las estadísticas de la organización mundial de la salud se registran 60 millones de infecciones asociadas a RSV por año, causando 200.000 muertes en todo el mundo [72-74]. La infección por RSV se manifiesta como una epidemia estacional asociada principalmente a los meses de invierno, o temporadas de lluvia según la región. Además de niños, esta infección afecta también a individuos inmunosuprimidos, inmunocomprometidos y ancianos. En Argentina, de los 350.000 casos de bronquiolitis reportados anualmente, el 60% son de origen viral, dentro de los cuales entre el 60 y el 80% de los aislamientos son RSV positivos [75]. A pesar del conocimiento acumulado sobre la biología celular y molecular del RSV, no hay vacunas ni tratamientos antivirales específicos eficientes para prevenir o tratar la infección [76].

Clasificación filogenética.

El RSV pertenece al orden *Mononegavirale*s. Los virus que forman parte de este orden se caracterizan por ser virus con cápsides helicoidales, envueltos por membranas y con genoma de ARN polaridad negativa (sentido 3'-5') no segmentado. Este orden incluye patógenos animales y humanos de gran importancia para la salud pública (Figura 10). Se encuentra conformado por 4 familias: Familia *Rhabdoviridae* (incluye patógenos humanos como el virus de la rabia), Familia *Bornaviridae* (incluye la enfermedad de Borna), Familia *Filoviridae* (incluye patógenos humanos como ebola y malburgo) y la Familia *Paramyxoviridae* (incluye patógenos humanos como el sarampión, el virus de la parotiditis humana, RSV, entre otros). Esta última puede subdividirse en dos subfamilias: la subfamilia *Paramyxovirinae* y la subfamilia *Pneumovirinae.* La subfamilia *Paramyxovirinae* esta conformada por 7 géneros e incluye patógenos humanos como Sarampión, virus de la parotiditis humana, virus de la parainfluenza humana tipo I-IV, virus Hendra y Nipha, entre otros. Por otra parte, la subfamilia *Pneumovirinae* esta compuesta por dos géneros: el género *Metaneumovirus* y género *Pneumovirus*, el cual incluye al RSV (Figura 10) [77]. El RSV existe como un solo serotipo que contiene a dos sub-grupos antigénicos A y B, los cuales se estiman que divergieron hace 350 años [77]. En particular, en esta tesis trabajaremos con proteínas del sub-grupo A2.

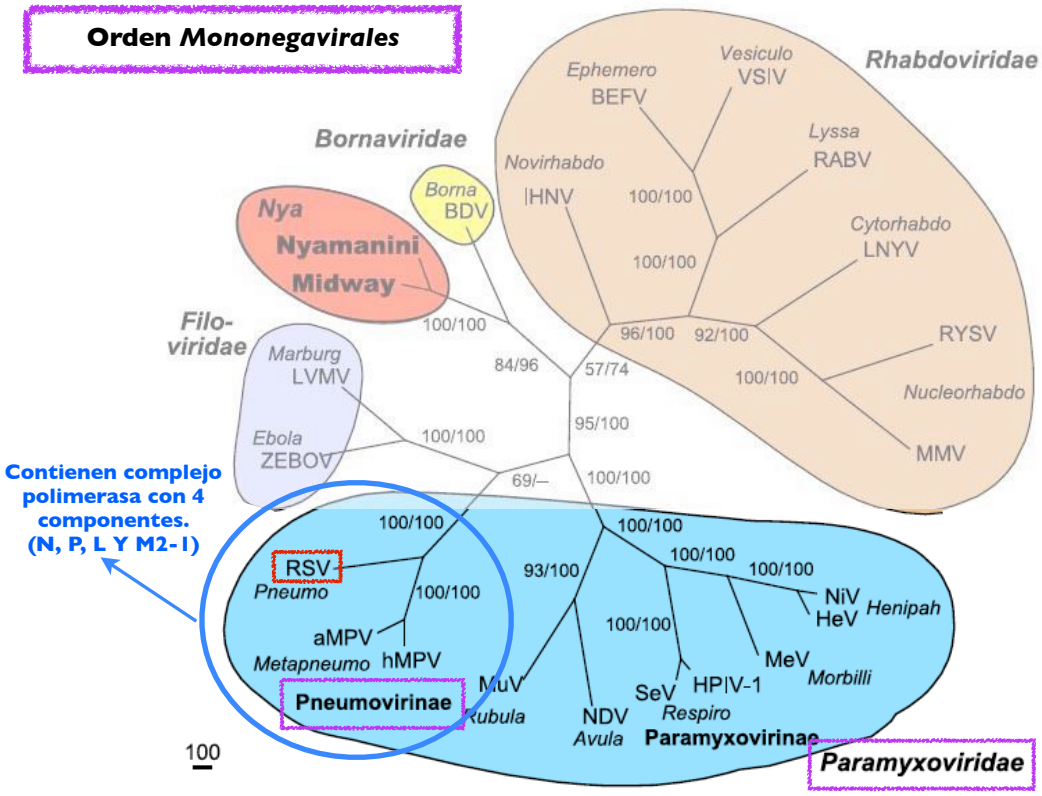


Figura 10. Arbol filogenético del órden *Mononegavirales*.

Características generales de RSV: genoma y proteínas.

El RSV posee uno de los mapas genómicos más complejos dentro de la familia *Paramyxoviridae*. Su genoma es de ARN simple cadena polaridad negativa no segmentado (aproximadamente 15.2 Kpb) (Figura 11), el cual presenta en los extremos 3´ y 5´ regiones no codificantes regulatorias llamadas *leader* y *trailer*, respectivamente [77]. Este genoma contiene 10 genes que codifican para 11 proteínas (Figura 11), de las cuales 6 (L, N, P, M, F, G) son comunes a todos los miembros del orden *Mononegavirales*. Además, 3 proteínas, 2 codificadas por el gen M2 (M_{2-1} y M_{2-2}) y la proteína SH, son comunes a todos los miembros de la subfamilia *Pneumovirinae* (Figura 10 y 11). Por otra parte, las proteínas no estructurales NS1 y NS2 son exclusivas de RSV y los miembros del género *Pneumovirus* [77]. En cuanto a su función, las proteínas F, G y SH son glicoproteinas que forman parte de la envoltura viral y median la interacción del virus con la célula huésped; la proteína de la matriz M es una proteína estructural que se asocia internamente con la membrana. Las proteínas N, P, L y M_{2-1} forman parte del complejo polimerasa viral; se cree que M_{2-2} regula la síntesis del ARN viral mediando el cambio entre la transcripción y la replicación. Además las proteínas no estructurales NS1 y NS2 participan en el bloqueo de la respuesta celular a interferón.

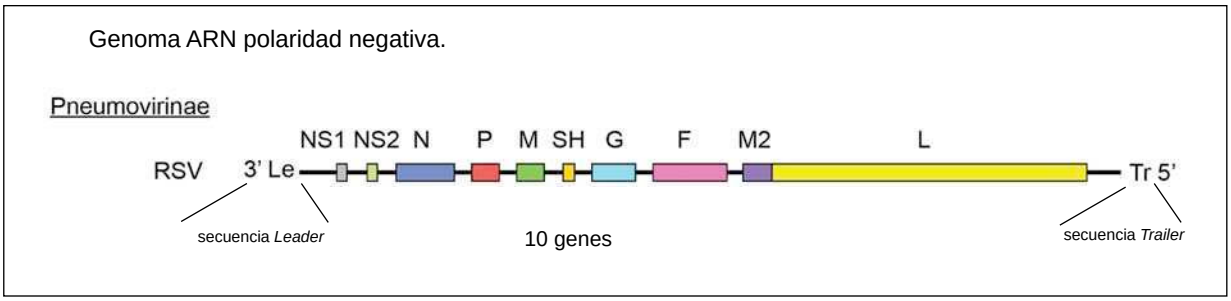


Figura 11. Organización genómica de RSV. Se indican los 10 genes que lo componen y su posición relativa en el genoma del virus, así como las secuencias *Leader* y *Trailer* regulatorias.

El RSV posee tropismo por células ciliadas del epitelio respiratorio. El virus ingresa a la célula a través de la fusión con la membrana plasmática, la cual es mediada por la glicoproteína F. Luego de la fusión, el material genético del virus asociado a la nucleocápside y al complejo polimerasa es liberado al citoplasma. El ARN polaridad negativa es usado como molde del complejo polimerasa viral para transcribir en forma secuencial los genes virales y generar ARNm subgenómicos, con *capping* y poliadenilados, a partir de los cuales son traducidas las proteínas virales. La replicación involucra la generación de una copia positiva de ARN complementaria (antigenoma), la cual sirve como molde para la síntesis del ARN genómico. La proteína M regula el ensamblado del virus mediante interacción con las glicoproteínas F y G, y con las proteínas del complejo polimerasa (N, P, L y M_{2-1}). La liberación de las partículas virales se da por brotación [77].

El complejo polimerasa de RSV.

El complejo polimerasa es un complejo proteico (en el rango de los megadaltons) esencial para la replicación y la transcripción del genoma de los virus del orden *Mononegavirales* durante la infección [78]. Según la familia viral, puede estar conformado por 3 o 4 proteínas: N, L y P, comunes a todos los integrantes del orden y la proteína M_{2-1} particular de la familia *Paramixoviridae*, subfamilia *Pneumovirinae* (Figura 10, círculo azul).

El complejo polimerasa de RSV se encuentra conformado por 4 proteínas: la proteína de la nucleocápside N que envuelve al ARN viral formando una nucleocápside helicoidal (NC), la polimerasa viral L, el cofactor de la polimerasa viral la fosfoproteína P y el factor antiterminador de la transcripción M_{2-1} (Figura 12) [77]. No se ha registrado la existencia de un complejo equivalente en la célula huésped, por lo tanto, el mismo representa un blanco atractivo para el desarrollo de antivirales específicos [79]. Sin embargo, el desarrollo de antivirales basados en la interferencia de sitios de interacción entre los distintos componentes del complejo no se han podido llevar a cabo hasta el momento debido que existe muy poca información estructural sobre las proteínas que lo componen, en particular de la polimerasa viral L y de su cofactor P.

A continuación se presenta una breve descripción de la información estructural conocida hasta el momento para los distintos componentes del complejo polimerasa de RSV:

- *Proteína de la Nucleocápside.* N es la proteína viral más abundante dentro de las células infectadas, la misma puede encontrarse en forma soluble y monomérica (N^0), o asociada al ARN viral formando parte de la nucleocápside (NC). En el 2009, el grupo del Dr. Felix Rey, reportó la estructura cristalina de un decámero de N en complejo con ARN [80] (Figura 12).

- *Factor anti-terminador de la transcripción M_{2-1}.* Esta proteína cumple la función de factor anti-terminador de la transcripción, la cual aumenta la procesividad de la polimerasa viral y previene la terminación prematura de los distintos transcriptos. Esta proteína es tetramérica en solución. Un grupo francés caracterizó la estructura del *core* monomérico de la proteína por RMN [81] y posteriormente la estructura cristalina del tetrámero [82] (Figura 12).

- *Polimerasa Viral L.* Es una proteína de aproximadamente 2000 aminoácidos, que posee todas las actividades enzimáticas requeridas para llevar a cabo todos los procesos involucrados en la replicación y transcripción viral, incluyendo: polimerización nucleotídica, *capping* y poliadenilación del ARNm, entre otros [78]. Su interacción con P es esencial tanto para estabilizar su estructura, como también para mediar la interacción con la NC (Figura 12) [78]. Hasta el momento, no se cuenta con información estructural de esta proteína o de otras del orden *Mononegavirales*, posiblemente debido a su gran tamaño y complejidad [78].

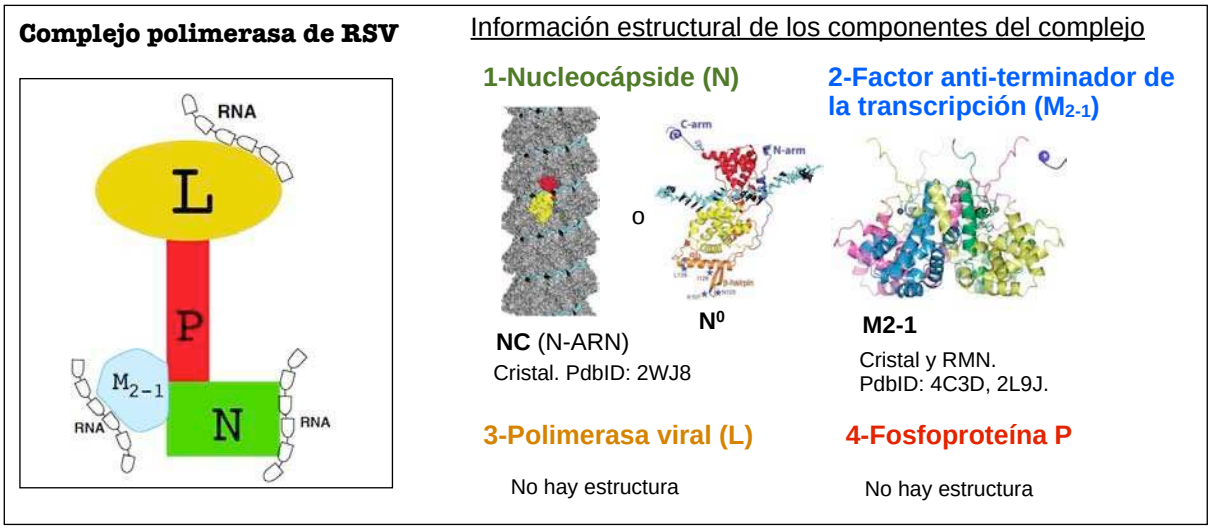


Figura 12. Representación esquemática del complejo polimerasa de RSV y detalle de las proteínas que lo conforman.

La fosfoproteína P.

La fosfoproteína P es un factor viral esencial para la transcripción y replicación de RSV, que no posee actividad enzimática asociada. Cumple una función central en el ensamblado del

complejo polimerasa, ya que media las interacciones entre los distintos componentes del mismo [78].

A continuación se listan algunas de sus principales funciones:

- *Co-factor de la polimerasa viral L.* Esencial para estabilizar y favorecer el reconocimiento entre L y la NC como templado.
- *Estabilización de los monómeros recién sintetizados de la proteína de nucleocápside (N^0)*. Bloquea la interacción inespecífica entre N^0 y el ARN.
- *Presentadora de N^0 como sustrato para la formación de la NC*.
- *Interacción con el factor antiterminador de la transcripción M_{2-1}.* Se presume que media la interacción entre M_{2-1} y el resto de los componentes de complejo.
- *Interacción con la proteína de la matríz (M).* Se propone que dicha interacción estaría involucrada en el desensamblado de la partícula viral.

Las fosfoproteínas P de las Familias *Paramixoviridae* y *Rhabdoviridae* presentan una organización modular conservada (Figura 13). Estas son proteínas extendidas que poseen un dominio de oligomerización central, el cual puede formar dímeros, trímeros o tetrámeros según el virus al que pertenezcan (Figura 13, cajas verdes). En general, los dominios de oligomerización están flanqueados por regiones con bajo contenido de estructura secundaria y terciaria, las cuales han sido predichos como intrínsecamente desordenadas (Figura 13, líneas rojas). El carácter IDP de las regiones predichas como desordenadas ha sido confirmado para los virus del sarampión (MeV), Nipha (NiV) y Hendra (HeV) [83], donde se describieron los comportamientos en solución de las regiones N-terminal y C-terminal utilizando diversas técnicas bioquímicas y biofísicas (Figura 13, rectángulos rojos).

Dentro de la subfamilia *Paramixovirinae* y de la familia *Rhabdoviridae* existe una gran cantidad de información sobre la estructura tridimensional de los distintos dominios estructurados de P. En particular, dentro de la región C-terminal, se describió un dominio con estructura globular tipo haz de tres hélices conocido como “dominio XD”, el cual interactúa con la región C terminal intrínsecamente desordenada de la NC viral (Figura 13, cajas azules) [83]. Además, recientemente se ha cristalizado para el paramixovirus Nipha [84] y para el rhabdovirus VSV [85] el complejo conformado por N^0 y los primeros residuos de la región IDP de P (residuos 1-50), donde esta última adopta una conformación α-hélice en el complejo, constituyendo un α-MORF (Figura 13).

Por otra parte, dentro de la subfamilia *Pneumovirinae,* a pesar de la importancia clínica de los patógenos que la conforman, existe escasa información estructural sobre las fosfoproteínas P. En particular, durante el desarrollo de esta tesis se publicó la primera estructura de un dominio de P de un miembro de dicha subfamilia, el dominio de tetramerización del metapneumovirus humano (hMPV) (Figura 13).

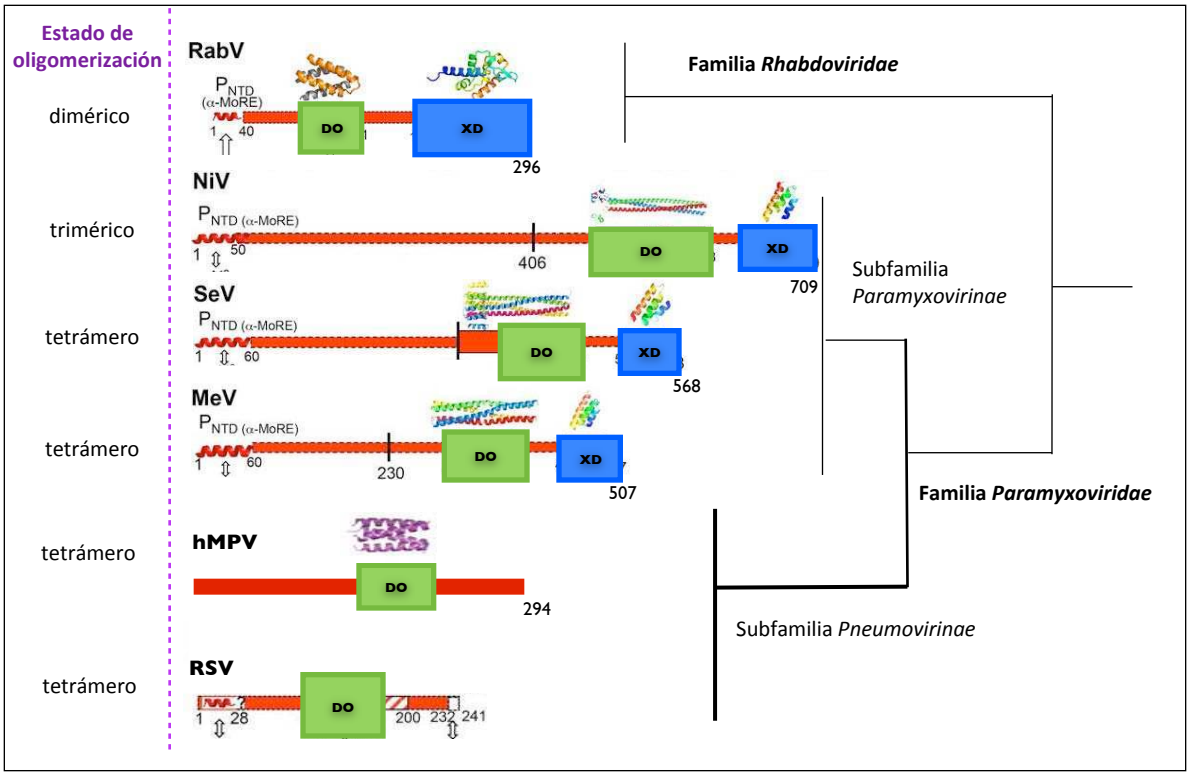


Figura 13. Esquema de la organización modular de P de las familias *Rhabdoviridae* y *Paramyxoviridae*. En rojo se muestran las regiones predichas como IDP, y las cajas verdes y azules representan al dominio de oligomerización y al dominio XD.

La fosfoproteína P de RSV es la variante más pequeña dentro de los miembros de la familia *Paramyxoviridae* (Figura 13). Es una proteína tetramérica en solución y esta conformada por 241 residuos. Hasta el momento, no se ha reportado estructura cristalográfica ni estructura por RMN de ninguna región de esta proteína, posiblemente debido a su elevada flexibilidad estructural. Mediante experimentos de proteólisis limitada con distintas proteasas se identificó un dominio central resistente (residuos 104-160), el cual contiene al dominio de tetramerización (residuos 119-160) [86] (Figura 14). Como única información estructural se cuenta con un modelo de la región central de P obtenido a partir de modelado por homología con el mismo dominio de P del virus Sendai (SeV) [87, 88] (Figura 14). Según este modelo, la región central de P está conformada por dos hélices, que en esta tesis llamaremos P_Z (residuos 104-119) y P_Y (residuos 119-160) (Figura 14). La hélice P_Y corresponde al dominio de tetramerización propiamente dicho y adopta estructura tipo *coiled-coil* conformada por un haz de cuatro hélices (Figura 14) [87].

Por otra parte, a partir de un estudio bioinformático realizado sobre la secuencia de P de RSV se describió, en homología con otros paramyxovirus, que la misma contiene una organización modular. Se definieron dos módulos: el módulo N-terminal (residuos 1-103) y el módulo C-terminal (residuos 161-241), los cuales fueron predichos mediante el uso de distintos algoritmos como intrínsecamente desordenados [86] (Figura 14).

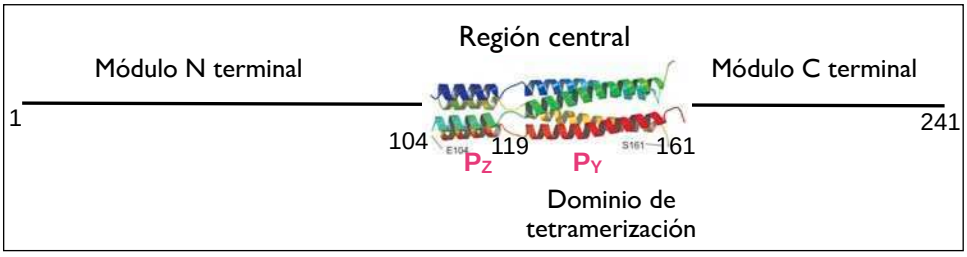


Figura 14. Representación esquemática de P de RSV. Se indica el modelo obtenido para la región central y la extensión de los módulos N y C terminal predichos a partir de experimentos de proteólisis, predicciones de desorden y similitud con otros Paramyxovirus.

Modificaciones post-traduccionales de P.

Las fosfoproteínas P de la familia *Paramyxoviridae* son fosforiladas en múltiples sitios durante el ciclo de vida del virus. Estas fosforilaciones, en su mayoría, se han descripto dentro de regiones identificadas como intrínsecamente desordenadas [83], sugiriendo que estas modificaciones post-traduccionales tienen impacto directo en la estabilización de elementos de estructura local.

En particular, P de RSV es fosforilada en varios residuos de serina y treonina, en la región N-terminal (S30, S39, S45, T46, S54) [89], en los residuos de la región central (T108, S116, 117, 119) [90, 91] y en el C-terminal (S232, S237) [92, 93] (Figura 15). Las fosforilaciones han sido asociadas a diversas funciones de P, como por ejemplo la S54 que media la interacción con M permitiendo el desensamblado de la partícula viral [94] o cumpliendo papeles regulatorios durante la transcripción del genoma [90, 95].

Interacciones con proteínas del complejo polimerasa.

Como ha sido mencionado anteriormente P puede interactuar con todos los componentes del complejo polimerasa (M_{2-1}, N, L) [78] y por lo tanto es considerada como la proteína que media el ensamblado del mismo [83]. Hasta el momento, se han identificado tres sitios de interacción en P de RSV, los cuales se localizan en las regiones predichas como IDPs (Figura 15, cuadros azules). Estos sitios fueron principalmente determinados mediante experimentos de inmunoprecipitación o *pull down* con distintas variantes mutadas o truncadas y validados en ensayos de función del complejo polimerasa en células. De esta forma, en el módulo N-terminal de P se identificó un sitio de interacción con N^0 (residuos 1-29) (Figura 15, cuadro azul) [96] y en el módulo C-terminal un sitio de interacción con la polimerasa viral L (residuos 212-239) [97] (Figura 15, cuadro azul). A partir del análisis con predictores de estructura secundaria y del efecto de mutaciones que alteran dicha estructura predicha, se propuso que ambos sitios podrían constituir elementos de reconocimiento molecular tipo α-MoRF [96, 97]. Por otra parte, el sitio de interacción con la NC se localiza en los últimos residuos de P (residuos 231-241) [98]. En un estudio reciente de la interacción NC-P se demostró por cristalografía de rayos X

que el complejo formado por la NC y los últimos 13 residuos de P, presenta características de complejos tipo *Fuzzy* [99].

Además, mediante técnicas de doble híbrido, se han identificado otros potenciales sitios de interacción en P (Figura 15, en naranja), sin embargo, la misma no permite distinguir si la interacción ocurre de forma indirecta o directa [100, 101] .

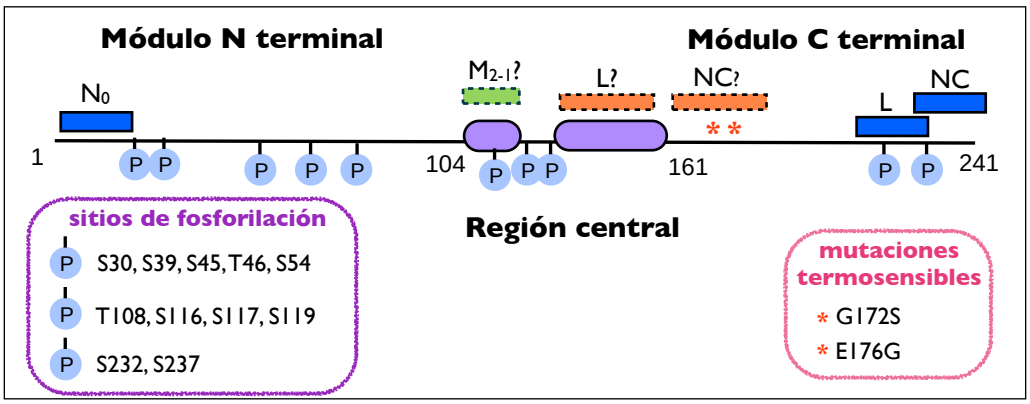


Figura 15. Representación esquemática de P, sus sitios de fosforlización y sitios de interacción con las proteínas del complejo polimerasa. En asteriscos rojos se indican las mutaciones termosensibles (G172S, E176G). En azul se indican los sitios de interacción localizados dentro de P por técnicas de pull down. Y en naranja se indican potenciales sitios implicados en interacciones identificados por doble híbrido y en verde un potencial sitio de interacción identificado por cromatografia de afinidad.

Por otra parte, se han identificado dos mutaciones termosensibles en el módulo C-terminal de P (G172S y E176G) que llevan a la pérdida de la interacción con la NC observado por doble híbrido como también por experimentos de inmunoprecipitación (Figura 15, asteriscos rojos) ([102]. En dicho trabajo, Lu *et al.* demostraron que ambas mutaciones cumplen un papel central en la replicación del virus, ya que los virus que las poseen no pueden replicar en cultivos celulares a temperatura fisiológica de 37 °C, pero si a 33 °C. Estas observaciones fueron validadas en experimentos de infección en ratas y ratones, observando el mismo comportamiento [102]. Además, demostraron que durante pasajes *in vitro* a 37 °C estas mutaciones revierten a un fenotipo que no presenta termosensibilidad [102].

RAZONAMIENTO, ALCANCE DEL TRABAJO Y OBJETIVOS.

Una de las grandes preguntas presentes en el campo de la virología, así como también una de las preguntas que motivó a este trabajo de tesis es: ¿Por qué los virus con tan pocos productos proteicos pueden interactuar de manera especifica con tantos blancos celulares? Como ya se introdujo, el desorden intrínseco en proteínas es una característica estructural asociada a favorecer la diversificación funcional de las proteínas. En particular, las proteínas intrínsecamente desordenadas o dominios desordenados están ampliamente distribuidos en proteomas virales, y actualmente la presencia de estos motivos, al igual que el *splicing* alternativo son reconocidos como características que brindarían ventajas adaptativas frente a la necesidad de utilizar eficientemente el material genético limitado que presentan.

El *objetivo general* de esta tesis fue aportar evidencias experimentales sobre IDPs virales a través de la caracterización bioquímica y biofísica de dos proteínas esenciales pertenecientes a dos modelos de virus de relevancia médica para nuestro país: la oncoproteína E7 de HPV-16 y el cofactor de la polimerasa viral la fosfoproteína P de RSV.

Esta tesis se dividió en dos capítulos:

-**Capítulo I.** ***Disección estructural del dominio E7N de HPV-16.***

En el primer capítulo, partiendo de la *hipótesis* de que existe una correlación entre las regiones con tendencia estructural local y los motivos lineales de interacción descriptos en E7N, se estudiaron los determinantes estructurales y equilibrios conformacionales de desorden intrínseco de la oncoproteína E7 de HPV-16. Este estudio permitió identificar tres elementos locales de estructura secundaria que existen en equilibrio con estados desplegados y los mismos contienen a todos los motivos lineales de interacción descriptos. Esta plasticidad conformacional presente en las regiones que contienen dichos motivos lineales de interacción permite dar una posible respuesta al gran número de blancos de interacción descriptos para este dominio.

-**Capítulo II.** ***Caracterización bioquímica y biofísica de la fosfoproteína P de RSV.***

Para este capítulo se desarrollaron y produjeron los distintos módulos y combinaciones de módulos de P en forma recombinante (Materiales y métodos). Utilizando las distintas proteínas recombinantes obtenidas, demostramos experimentalmente por primera vez la naturaleza IDP de la proteína P de RSV. La misma contiene dos módulos intrínsecamente desordenados, los cuales presentan distinto contenido de desorden. Definimos al dominio C-terminal de P como un dominio con características en solución de tipo *pre-molten globule like IDP*, y al dominio N-terminal como un dominio con características de tipo *random-coil like IDP*. Identificamos que P

presenta un comportamiento térmico anómalo, presentando en el dominio C-terminal un módulo de estructura modulable por pequeñas variaciones de temperatura en el rango de temperatura fisiológico.

RESULTADOS

CAPITULO I

Disección estructural del dominio E7N de HPV-16.

DISECCIÓN ESTRUCTURAL DEL DOMINIO E7N DE HPV-16.

En un trabajo previo de nuestro laboratorio, se demostró que los primeros 40 residuos de la oncoproteína E7 de HPV-16 (E7N) conforman un domino *per se*, que puede ser definido como intrínsecamente desordenado. Se describió que este dominio, a pesar de carecer estructura secundaria definida en solución, puede estabilizar estructuras tipo α-hélice, PII y lámina β bajo distintas condiciones de solvente y pH [64]. Dado que esta región se encuentra densamente poblada por motivos lineales de interacción, hipotetizamos que dichas transiciones conformacionales afectarían la exposición de los mismos.

Una característica de las IDPs es que pueden estabilizar localmente elementos de estructura secundaria transitoria, los cuales pueden experimentar transiciones entre distintas conformaciones. En base a estas características, nos preguntamos: ¿Son aquellas estructuras estabilizadas en E7N [64] de naturaleza local, o requieren la información de secuencia completa del dominio para su estabilización? Suponiendo que existen elementos de estructura local dentro de E7N, ¿Están dichos elementos asociados a los motivos lineales de interacción?

Para dar respuesta a estos interrogantes, nos propusimos realizar una disección estructural de los elementos de estructura secundaria que pueden ser estabilizados dentro del domino E7N. En primer lugar utilizamos péptidos sintéticos correspondientes a distintos fragmentos del dominio E7N. Se determinaron sus propiedades hidrodinámicas mediante las técnicas de RMN y ultracentrifugación analítica. Se investigó la presencia de elementos de estructura secundaria local utilizando CD al UV-lejano y se analizó la presencia de estructura pre-existente a través de estabilización por solventes. Por otra parte, se asignó por RMN el péptido correspondiente al dominio E7N completo y se determinó a nivel atómico la posición de los elementos de estructura secundaria local estabilizados dentro del mismo.

I.1.- CARACTERIZACIÓN DE LOS ELEMENTOS DE ESTRUCTURA PRESENTES EN E7N POR FRAGMENTACIÓN Y ESTABILIZACIÓN POR SOLVENTES.

Para poder identificar los elementos estructurales estabilizados dentro del domino IDP E7N, diseñamos una serie de péptidos sintéticos abarcando distintas regiones del mismo. En la figura I.1, se muestran los detalles de los distintos péptidos estudiados. Brevemente:

- **E7(1-40)** comprende todo el domino E7N incluyendo las regiones conservadas CR1 y CR2 y los motivos lineales de interacción identificados en este dominio.
- **E7(1-20)** incluye toda la región conservada CR1, y los motivos lineales E2F-mimic y de fosforilación para la kinasa DYRK1A.

- **E7(16-40)** abarca toda la región conservada CR2 incluyendo a los motivos lineales LxCxE, el motivo de fosforilación por la enzima caseina kinasa II (CKII), y un motivo de degradación PEST.
- **E7(16-31)** comprende parte de la región conservada CR2 pero excluye la región que contiene a los motivos CKII y PEST.
- **E7(21-29)** contiene únicamente al motivo LxCxE.
- **E7(25-40)** contiene únicamente a los motivos CKII y PEST.

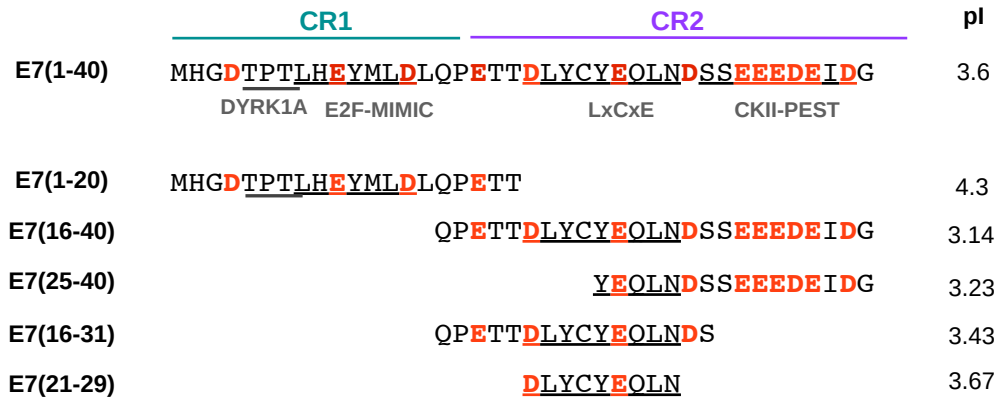


Figura I.1. Esquemas de los fragmentos utilizados. Se detallan las posiciones de las regiones conservadas CR1 y CR2, así como también la ubicación de los motivos lineales contenidos en los distintos fragmentos. A la derecha de las secuencias se indican los puntos isoeléctricos (pI) de cada secuencia y en rojo se indican los residuos ácidos.

Dado que la Cys 24 del dominio E7N presenta una fuerte tendencia a oxidarse formando uniones de tipo puente disulfuro entre monómeros, todos los resultados presentados en este capitulo fueron realizados en condiciones reductoras. Los experimentos de CD fueron realizados en 1 mM DTT concentración final en buffer, mientras que los experimentos de RMN fueron realizados en 10 mM TCEP.

Una característica distintiva del dominio E7N, y por lo tanto de varios de los fragmentos que lo componen, es la elevada proporción de residuos ácidos en su secuencia (Figura I.1). Estas cargas negativas del dominio E7N (pI 3.6) pueden ser neutralizadas en condiciones de pH bajo, llevando a un plegamiento parcial del dominio. En un trabajo previo de nuestro laboratorio, se demostró que el dominio E7N puede estabilizar estructura α-hélice y que dicha estabilización estructural es favorecida a bajo pH [64], indicando que la neutralización de los carboxilatos podría estar favoreciendo la compactación del dominio.

I.1.1- Efecto del pH sobre el radio hidrodinámico del dominio E7N determinado por RMN.

Una característica de las proteínas IDPs es su comportamiento hidrodinámico anómalo, manifestado en radios de Stokes (R_S) intermedios entre un estado completamente desplegado y su contraparte globular. Para determinar los radios de Stokes (R_S) para el dominio E7N en

distintas condiciones de protonación, se realizaron experimentos de RMN de gradiente de campo pulsado (PFG-NMR, por sus siglas en inglés) a pH 7.5 y pH 5.0. Los experimentos de RMN presentados en esta tesis fueron realizados por la Dra. Mariana Gallo del laboratorio de RMN de la Fundación Instituto Leloir, en colaboración con el presente proyecto.

Las PFG-NMR son medidas de difusión que consisten en someter a la muestra a un gradiente de campo magnético en el eje z, paralelo al campo magnético estático B_0, dejar un periodo en el cual la molécula difunde y después aplicar el mismo gradiente con el signo cambiado. Se analiza entonces la variación de la intensidad de los espectros monodimensionales a medida que aumenta la fuerza del gradiente [34] (Ver materiales y métodos).

Los R_S determinados para el dominio E7N a pH 7.5 y pH 5.0 fueron de 15.3 ± 0.2 Å y 14.3 ± 0.3 Å, respectivamente, demostrando la compactación del dominio al disminuir el pH. Esta compactación en E7N puede explicarse por la protonación parcial de las cargas de los residuos glutámicos y aspárticos a pH 5.0, las cuales disminuyen la repulsión electrostática de las cadenas laterales de los mismos, favoreciendo la estabilización de una conformación más compacta.

Por otra parte, comparamos los R_S determinados para E7N con los valores teóricos esperados para una proteína de 40 residuos en una conformación globular y una conformación desplegada en condiciones desnaturalizantes. Para ello se utilizaron las relaciones empíricas definidas por Wilkins *et. al.* a partir de experimentos de PFG-NMR [34], las cuales permiten determinar los R_S esperados teóricos en estado globular ([Ecuación 1]) o en estado desplegado ([Ecuación 2]) a partir del número de residuos de la cadena polipeptidica (N):

[Ecuación 1] $R_S = 4.75\ N^{0.29}$ Å

[Ecuación 2] $R_S = 2.21\ N^{0.57}$ Å

Los R_S teóricos esperados para un péptido de 40 residuos en conformación globular (estado teórico) o totalmente desplegado en desnaturalizante son de 13.8 o 18.1 Å, respectivamente [34]. Los valores obtenidos experimentalmente de 15.3 ± 0.2 Å a pH 7.5 y 14.3 ± 0.3 Å a pH 5.0, resultan intermedios entre los esperados para ambas conformaciones. Esto indica que E7N posee cierto grado de compactación en solución tanto a pH 7.5 como a pH 5.0, ya que en ambos casos presenta radios hidrodinámicos menores que los esperados para un dominio completamente desplegado. Por lo tanto, estos resultados iniciales sugieren que dentro del conjunto de posibles conformaciones que puede estabilizar E7N, este dominio puede estar generando contactos terciarios y/o poblando elementos de estructura secundaria transitorios que permitan explicar la compactación parcial observada para el mismo.

I.1.2- Análisis de las propiedades hidrodinámicas de E7N por AUC.

Para determinar las propiedades hidrodinámicas del dominio E7N y de los fragmentos E7(1-20) y E7(16-40) correspondientes a las regiones conservadas CR1 y CR2, realizamos medidas de ultracentrifugación analítica mediante el método de velocidad de sedimentación (AUC-SV, por sus siglas en inglés). La misma consiste en someter a las muestras a centrifugación y medir la distribución de la concentración de las partículas en la celda durante la migración de las macromoléculas. Como resultado se obtiene el coeficiente de sedimentación *(s)* a partir del cual se pueden determinar los pesos moleculares y los parámetros relacionados a la forma de las partículas. Estos experimentos fueron realizados en colaboración con el Dr. Andrés Salvay de la Universidad Nacional de La Plata. Medidos en Grenoble en el laboratorio de la Dra. Christine Ebel.

Los perfiles de sedimentación fueron determinados a tres concentraciones distintas para los tres péptidos (Figura apéndice A11). En la figura I.2-A se muestra la superposición de los perfiles de sedimentación obtenidos para E7N, E7(1-20) y E7(16-40) a 1mg/ml.

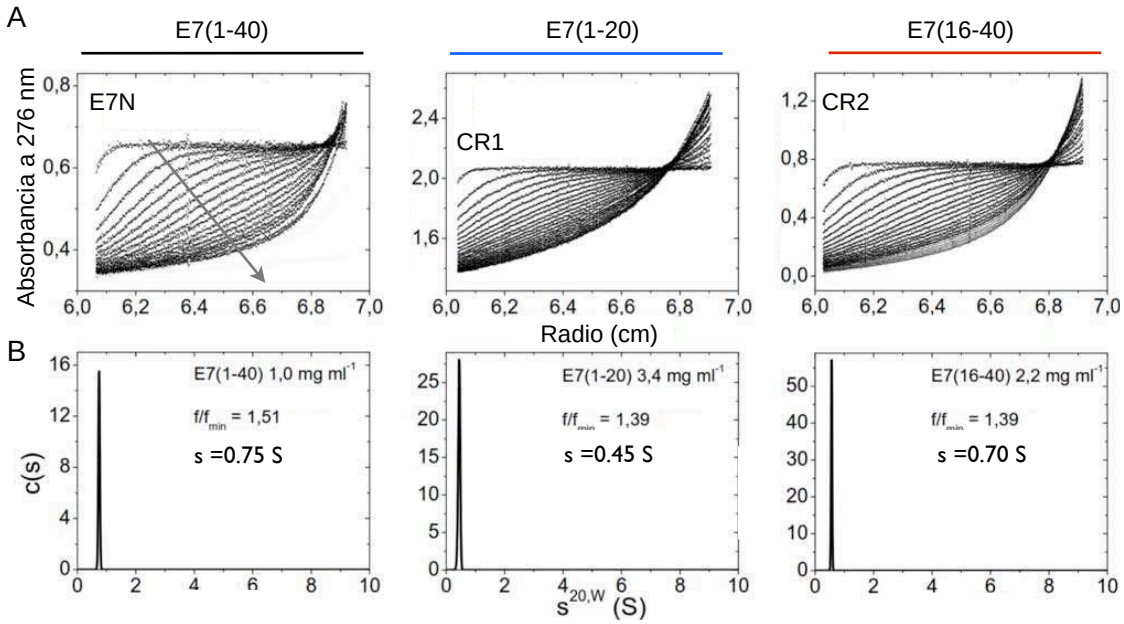


Figura I.2. Medidas de velocidad de sedimentación por AUC para E7N, CR1 y CR2. A-Superposición de los perfiles de sedimentación experimentales , medidos a 1.0, 3.4 y 2.2 mg/ml para E7(1-40), E7(1-20) y E7(16-40), respectivamente. La última medida corresponde a un perfil de sedimentación de 24 horas. B-Distribución c(s) en el rango 0 a 10 S. La señal fue normalizada a 1 cm de paso óptico. Todas las medidas de AUC fueron realizadas en 50 mM Tris-HCl (pH 7.5) y 1 mM DTT, 50.000 rpm y 20 ºC. 1 Svedberg= 10^{-13} segundos. La flecha indica el sentido de la determinación.

Cada curva representa un perfil de sedimentación individual determinado por la absorbancia a 276 nm a lo largo del radio de la celda (cm) registrados a intervalos de 13 minutos durante 24 horas. Los mismos fueron analizados en términos de un modelo de distribución continua de partículas *c(s)* (Figura I.2-B). Se puede observar que todos los péptidos presentaron perfiles de sedimentación homogéneos dentro del rango entre 0 y 10 Svedbergs (S, donde 1S= 10^{-13} seg), correspondientes a la presencia de un solo tipo de especie mayoritario y bien definido. Dichos

perfiles homogéneos fueron obtenidos a las tres concentraciones estudiadas (Figura apéndice A11).

Por otra parte, los perfiles de velocidad de sedimentación fueron analizados en términos de especies independientes no interactuantes para poder estimar el R_S y el peso molecular aparente (MW_{app}) de cada uno de los fragmentos. Para ello se analizaron los coeficientes de difusión o la inversa de los coeficientes de sedimentación en función de la concentración de los distintos fragmentos. De los datos extrapolados de las curvas, se determinaron los coeficientes de sedimentación s_0 (Figura I.3-A) y los coeficientes de difusión D_{0app} (Figura I.3-B) en dilución infinita. Dichos valores fueron utilizados para calcular los valores de MW_{app} y el R_S para cada fragmento (Tabla I.1) (Ver materiales y métodos).

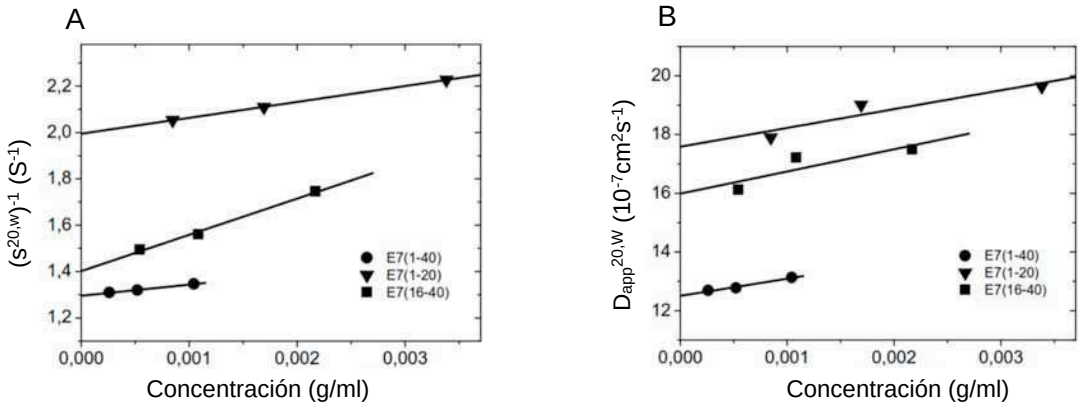


Figura I.3. Dependencia de los coeficientes de sedimentación y de difusión con la concentración de los distintos péptidos. A-Análisis de la dependencia de la inversa del coeficiente de sedimentación (s^{-1}) con la concentración. B-Análisis de la dependencia del coeficiente de difusión (D) con la concentración de muestra. De la regresión lineal de las curvas se obtienen los valores de s_0 y de D_{app0}.

Todos los MW_{app} calculados resultaron similares a los pesos moleculares teóricos obtenidos a partir de la secuencia primaria de la proteína utilizando el programa Expasy protParam (MW_{Expasy}) (Tabla I.1), indicando que E7N y los fragmentos correspondientes a las regiones CR1 y CR2 son monoméricos en solución. Por otra parte, los R_S determinados para E7N y los fragmentos E7(1-20) y E7(16-40), fueron de 17 ± 1 Å, 12 ± 1 Å y 13 ± 1 Å, respectivamente.

Tabla I.1. Propiedades hidrodinámicas de E7N, CR1 y CR2.

Fragmento	$S_0^{20,H2O}$ (S)	$D_{0app}^{20,H2O}$ (10^{-7} cm^2 s^{-1})	MW_{app} (KDa)	MW_{Expasy} (KDa)	f/f_{min}	R_S AUC (Å)	R_S Globular (Å)	R_S Desplegado (Å)
E7N (E7(1-40))	0.77 ± 0.01	12.5 ± 0.1	4.9 ± 0.2	4.68	1.53 ± 0.01	17 ± 1	13	19
CR1 (E7(1-20))	0.50 ± 0.01	17.6 ± 0.5	2.4 ± 0.2	2.33	1.40 ± 0.02	12 ± 1	10	13
CR2 (E7(16-40))	0.71 ± 0.01	15.9 ± 0.6	3.4 ± 0.2	2.92	1.45 ± 0.03	13 ± 1	11	14

Para poder comparar los R_S calculados por AUC con los valores teóricos esperados para especies globulares o 100% desplegadas en desnaturalizante (8M Gdm.Cl), se calcularon los

R_S teóricos para los distintos péptidos utilizando un modelo propuesto por Uversky [103]. Dicho modelo relaciona el peso molecular (MW, por sus siglas en inglés) de una proteína o péptido con los R_S en distintas condiciones. A continuación se presentan las ecuaciones derivadas de dicho modelo, las cuales fueron utilizadas para determinar los R_S teóricos esperados para una conformación globular [Ecuación 3] o para una conformación 100% desplegada en Gdm.Cl [Ecuación 4], [103]:

[Ecuación 3] $\log(R_S) = -(0.204 \pm 0.023) + (0.357 \pm 0.005) \log(MW)$

[Ecuación 4] $\log(R_S) = -(0.723 \pm 0.033) + (0.543 \pm 0.007) \log(MW)$

donde los MW están expresados en Da y los R_S en Anstromgs.

Podemos observar que el dominio E7N presenta R_S intermedios a los esperados para una proteína desplegada y su contraparte globular a pH 7.5 (Tabla I.1), similar a lo observado en las determinaciones por PFG-NMR. Al comparar los R_S determinados a pH 7.5 por PFG-NMR y por AUC para el domino E7N, encontramos diferencias (17 ± 1 Å y 15.3 ± 0.2 Å), posiblemente asociadas a las diferencias intrínsecas entre las técnicas utilizadas para las determinaciones.

Por otra parte, los R_S de los fragmentos correspondientes a las regiones CR1 y CR2 poseen un valor absoluto intermedio entre los esperados para una especie globular y una especie desplegada en Gdm.Cl. Sin embargo, teniendo en cuenta el error de la medida de AUC, los R_S para CR1 y CR2 coinciden con los R_S esperados para estados desplegados (Tabla I.1). Lo cual sugiere que las regiones CR1 y CR2 a pH 7.5 se encuentran mayormente desplegados cuando no están en el contexto del dominio E7N.

Asimismo, estimamos el cociente de fricción (f/f_{min}) para los distintos péptidos (Ver materiales y métodos), el cual permite describir la forma de las macromoléculas en solución, ya que depende de la hidratación, la forma, y las irregularidades de la partícula [33]. Para macromoléculas globulares y compactas con pesos moleculares de entre 5 y 1000 KDa, este parámetro varía entre 1.15 y 1.3, respectivamente [33]. Sin embargo, se ha demostrado para IDPs, que los valores de f/f_{min}, son significativamente mayores a los reportados para proteínas globulares, adoptando valores de f/f_{min} entre 1.5 a 3 para IDPs entre 5 y 200 KDa [33]. Los valores obtenidos de f/f_{min} para E7N y los péptidos E7(1-20) y E7(16-40) fueron de 1.53, 1.4 y 1.45, respectivamente. Dichos valores analizados en conjunto con los pesos moleculares de cada fragmento, los cuales no superan los 5 KDa, confirman que E7N y los distintos fragmentos presentan características hidrodinámicas extendidas típicas de IDPs (Tabla I.1).

En su conjunto, los resultados de AUC-SV y PFG-NMR indican que el dominio E7N posee características hidrodinámicas extendidas típicas de IDPs presentando cierto grado de compactación en las condiciones estudiadas. Dicho dominio posee un R_S intermedio entre una especie globular y una especie desplegada en desnaturalizante, mientras que los fragmentos

correspondientes a la región CR1 (E7(1-20)) y a la región CR2 (E7(16-40)), presentan R_S similares a los esperados para estados desplegados. Estos resultados sugieren la existencia de posibles contactos transitorios entre las regiones CR1 y CR2.

I.1.3- Análisis del contenido de estructura secundaria local por dicroísmo circular.

El CD en el UV lejano, como fue explicado en la introducción, es una técnica que permite evaluar el contenido de estructura secundaria de un polipéptido debido a que proporciona información sobre la estructura que adopta la cadena peptídica del mismo. La inspección de la forma del espectro y la posición de las bandas correspondientes a máximos y mínimos característicos permiten obtener información referida al tipo de estructura secundaria presente en el mismo (ver introducción).

La señal de CD al UV lejano a una determinada longitud de onda puede ser interpretada como una combinación lineal entre los aportes de los distintos componentes de estructura secundaria:

$$\text{Señal} = f_\alpha[\Theta]_\alpha + f_{PII}[\Theta]_{PII} + f_\beta[\Theta]_\beta + f_C[\Theta]_C$$

donde f_α, f_{PII}, f_β y f_C corresponden a la fracción de estructura tipo α-hélice, PII, lámina-β y *coil*, respectivamente. Por otra parte, $[\Theta]_\alpha$, $[\Theta]_{PII}$, $[\Theta]_\beta$ y $[\Theta]_C$, corresponden a los valores de las elipticidades molares en dichas conformaciones. La determinación del porcentaje exacto de las distintas conformaciones α, β, PII o *coil* presente en un polipeptido no es sencillo de definir, debido a la superposición de señales de los distintos tipos de estructura a una determinada longitud de onda y a las diferencias en sus coeficientes de absorción.

Con el fin de analizar el contenido de estructura secundaria presente en los distintos fragmentos de E7N, se midieron espectros de CD al UV lejano para cada uno de ellos en buffer a pH 7.5 y a 20 ℃. Todos los fragmentos estudiados presentan espectros de CD con características generales de polipéptidos desordenados, con un mínimo alrededor de 198 nm y ausencia de señal de CD a 220 nm (Figura I.4-A y B). Se puede apreciar que la mayoría de los fragmentos presentan valores de elipticidad a 200 nm similares (-15.000 deg.cm^2.dmol^{-1}) (Figura I.4-A). Sin embargo, los subfragmentos de la región CR2, E7(21-29) y E7(25-40), presentan espectros de CD que difieren sustancialmente del resto (Figura I.4-B). Como se puede observar en la figura I.4-B, el espectro de CD del fragmento E7(21-29) (línea naranja y triángulo azul) presenta una banda positiva a 230 nm. Dicha banda positiva puede ser interpretada inicialmente como una conformación secundaria tipo *turn* o giro. Sin embargo, dado el porcentaje alto de residuos aromáticos de tirosina (aproximadamente 20%) que presenta este fragmento, no podemos descartar que dicha banda positiva se deba al aporte de los mismos a la señal de CD. Un efecto similar fue observado para el péptido LH2-CP de la

proteína *barstar* el cual contiene 18% de residuos aromáticos y presenta una banda positiva a 230 nm, similar a la observada para E7(21-29) [104].

Por otro lado, el péptido E7(25-40) presenta un mínimo a 200 nm con valores de elipticidad menores en comparación al resto de los fragmentos (-22.000 deg.cm^2.dmol^{-1} *versus* -15.000 deg.cm^2.dmol^{-1}) (Figura I.4-B, línea magenta y triángulo rojo). Dada la forma del espectro y el elevado contenido de residuos ácidos (aproximadamente 50%) presentes en dicho fragmento (Figura I.1), sugerimos que estos cambios en el espectro de CD estarían reportando la presencia de estructura tipo PII.

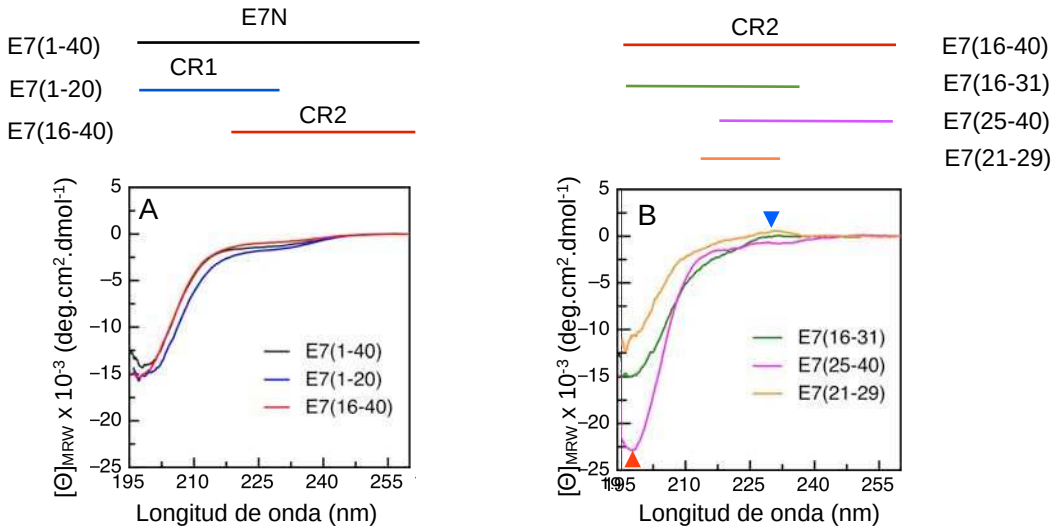


Figura I.4. Espectros de CD al UV-lejano de los distintos fragmentos pertenecientes a E7N. A-Espectros de CD de E7N (línea negra) y los fragmentos correspondientes a las regiones conservadas CR1 (línea azul) y CR2 (línea roja). B-Espectros de CD de E7(16-31) (línea verde), E7(25-40) (línea magenta) y E7(21-29) (línea naranja). Todos los espectros de CD fueron medidos en buffer 10 mM Tris-HCl pH 7.5 y 1 mM DTT, a 20℃ a una concentración de péptidos de 20 a 40 µM. En la parte superior de los gráficos se esquematiza un mapa de los fragmentos utilizados. El triángulo rojo y azul indican las diferencias entre espectros.

Estos resultados sugieren que el dominio E7N, a pesar de carecer de estructura secundaria consolidada en las condiciones estudiadas, posee elementos de estructura transitoria local, los cuales determinan las diferencias observadas en los espectros de CD de los distintos fragmentos.

I.1.4- Predicción de tendencia α-hélice dentro del dominio E7N.

En primer lugar, evaluamos la tendencia α-hélice dentro del dominio E7N mediante el uso del algoritmo AGADIR [105]. El mismo predice la tendencia local de una secuencia aminoacídica a adoptar estructura α-hélice en ausencia de un plegamiento terciario [105]. En la figura I.5 se muestra la predicción de estructura obtenida para E7N expresada como tendencia hélice en función del número de residuos del dominio. Podemos identificar tres regiones con la capacidad de estabilizar hélices dentro de E7N. Una primera hélice predicha en la región CR1 (residuos

P6-Q16) y otras dos hélices en la región CR2, las cuales comprenderían los residuos D21-D30 y S32-D39 (Figura I.5).

Como se mencionó anteriormente, las regiones CR1 de E7 y de la proteína E1A de adenovirus (AdE1A) presentan una elevada identidad de secuencia y funcional. Estudios cristalográficos revelaron que la región CR1 de AdE1A adopta conformación α-hélice en el complejo con pRb [106]. Por lo tanto, la presencia de una hélice en la región CR1 podría estar indicando un mecanismo de interacción similar a AdE1A. Por otra parte, es interesante destacar que en la región CR2 se predice una modulación de la tendencia a estabilizar hélice por el pH (Figura I.5, flecha negra).

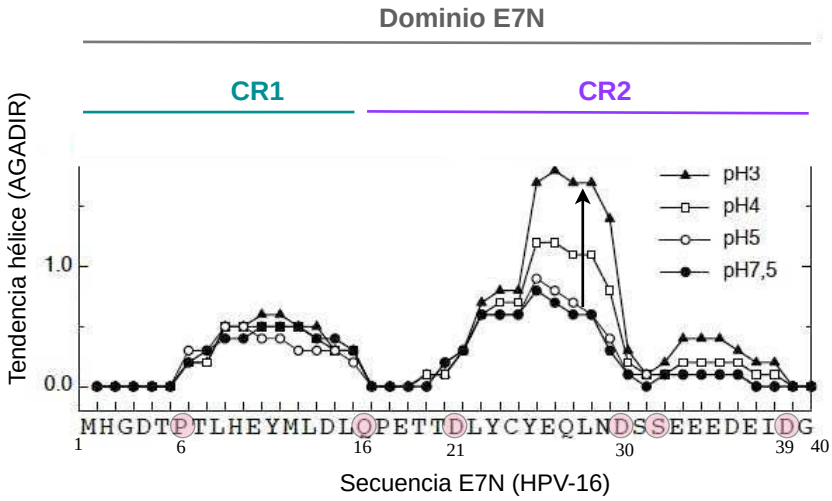


Figura I.5. Predicción de la tendencia local a la formación de α-hélice en el dominio E7N usando el programa AGADIR. Las medidas fueron predichas a distintos pHs. En la parte superior de la figura se indica la extensión de las regiones conservadas CR1 y CR2.

Las proteínas intrínsecamente desordenadas se caracterizan por presentar baja hidrofobicidad y elevada carga neta. Existen numerosos ejemplos en literatura que demuestran que pequeñas variaciones en el entorno físico-químico que lleven a un aumento en la hidrofobicidad del medio o una reducción en la carga neta de la proteína, promueven cambios en regiones con tendencias estructurales locales dentro de las IDPs que representan conformaciones fisiológicas posibles [23]. A continuación se abordará el análisis de los distintos componentes de estructura secundaria local de E7N estabilizados por el agregado de distintos co-solventes.

I.1.5- Estabilización de estructura α-hélice por 2,2,2-trifluoroetanol (TFE).

Uno de los solventes más utilizados para la estabilización de equilibrios pre-existentes entre regiones con tendencia local α-hélice y sus formas desordenadas es el TFE [107]. Dicho co-solvente estabiliza elementos pre-existentes de estructura α-hélice, desplazando el equilibrio hacia la población con dicha estructura conformada [107]. Con el fin de determinar la presencia de α-hélices locales y transitorias dentro del dominio E7N, nos propusimos evaluar el efecto del TFE sobre los distintos fragmentos del mismo. Para ello se realizaron titulaciones con TFE

seguidas por CD al UV lejano. Dada la abundancia de cargas negativas con posible participación en la modulación de la conformación de los fragmentos, sumado a los resultados obtenidos acerca del impacto del pH en el R_S de E7N, decidimos evaluar la influencia del pH en estas transiciones en condiciones de buffer a pH 7.5 y pH 5.0.

Podemos observar para E7N (E7(1-40)) y para los fragmentos E7(1-20), E7(16-40) y E7(16-31) que el agregado de concentraciones crecientes de TFE produce estabilización de estructura tipo α-hélice a ambos valores de pH (Figura I.6-A, columna 1 a 4). Esto se aprecia por el corrimiento del mínimo espectral de 198 nm hacia 208 nm y por el concomitante aumento de la señal correspondiente a la banda a 222 nm (Figura I.6-A, columna 1 a 4). Además, se puede observar que todas las transiciones se cruzan en un punto a aproximandamente 203 nm. Dicho punto se denomina isodicroico y es característico de transiciones entre dos estados, en este caso entre *coil* y α-hélice (Figura I.6-A, columna 1 a 4) [104].

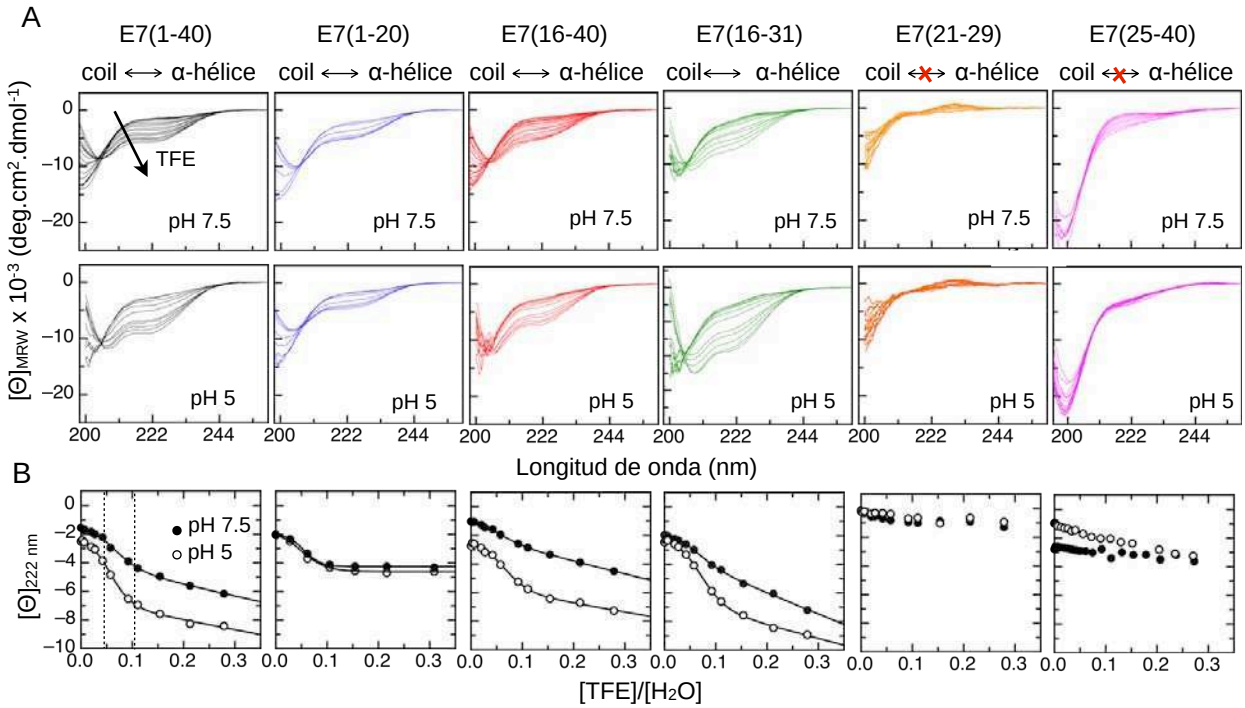


Figura I.6. Titulaciones de los distintos fragmentos de E7N con TFE a pH 7.5 y pH 5.0. A-Espectros de CD en el UV lejano de los distintos péptidos de E7N en mezclas de TFE-H_20 cubriendo el rango 0-50% TFE v/v a pH 7.5 (panel superior) y a pH 5.0 (panel inferior). B-Titulaciones con TFE seguidas por la variación de la elipticidad molar de la banda a 222 nm a diferentes relaciones [TFE]/[H_2O] a pH 7.5 (círculos negros) y a pH 5.0 (círculos blancos). Los datos fueron ajustados a un modelo de equilibrio de dos estados coil-hélice. Para los péptidos E7(21-29) y E7(25-40) los datos no pudieron ser ajustados debido a la ausencia de transición. La flecha indica el sentido de la titulación. La línea punteada indica la relación molar correspondientes a 12 y a 30% TFE, respectivamente. Todos los espectros fueron medidos en buffer 10 mM Tris-HCl pH 7.5, o en buffer 10 mM formiato de sodio pH 5.0, a 20℃ a una concentración de péptidos de 40 a 80 µM.

Por otra parte, para los fragmentos contenidos en la región CR2; E7(21-29) y E7(25-40), si bien el agregado de TFE induce un cambio espectral, el aspecto de los mismos no es característico de α-hélice (Figura I.6-A, columnas 5 y 6). En particular, el fragmento E7(21-29) está contenido en el mínimo fragmento del CR2 que estabiliza hélice (E7(16-31)), lo cual indica que la

información contenida en el mismo no alcanza para estabilizar dicha hélice. A su vez, para el fragmento E7(25-40) se puede observar que el agregado de TFE lleva a la disminución en la intensidad del mínimo a 200 nm con un ligero aumento en valor absoluto de la banda a 220 nm (Figura I.6-A, columna 6). La forma y la característica lineal no cooperativa de la transición sugiere que el agregado de TFE estaría desestabilizando posibles estructuras PII adoptadas por este fragmento (Figura I.6-A, columna 6). En conclusión, los resultados mostrados hasta ahora sugieren que existen al menos dos regiones con la capacidad de estabilizar hélice en presencia de TFE dentro de E7N. Una localizada en la región CR1 (péptido E7(1-20), y otra localizada en la región CR2 (contenida en el fragmento E7(16-31).

Para evaluar la magnitud de la estabilización de estructura α-hélice en los distintos péptidos del dominio E7N, se analizaron las transiciones estructurales obtenidas con el agregado de concentraciones crecientes de TFE, mediante el análisis del cambio de la señal de la banda a 222 nm (Figura I.6-B). En primer lugar, podemos observar que todos los fragmentos, con excepción del péptido E7(1-20), pueden estabilizar una mayor proporción de estructura α-hélice al bajar el pH (Figura I.6-B). Asimismo, al inspeccionar la forma de las transiciones podemos observar que para el dominio E7N y los fragmentos E7(16-40) y E7(16-31) se observa una mayor cooperatividad en la transición a pH 5.0 que a pH 7.5, en comparación con E7(1-20) (Figura I.6-B, columna 1 a 4).

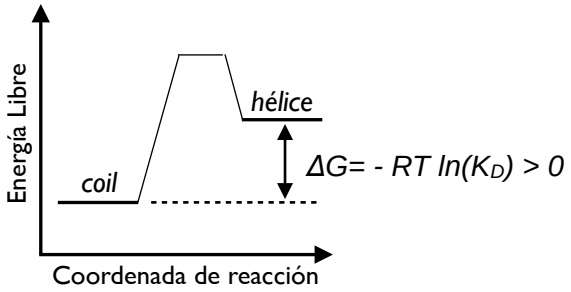
Los datos de las titulaciones fueron ajustadas a un modelo que propone la existencia de un equilibrio entre dos estados, un estado *random coil* y un estado α-hélice [108]:

$$\textit{coil} \rightleftharpoons \alpha\text{-hélice}$$

Donde podemos expresar a la K_D como:

$$K_D = \frac{[\alpha\text{-hélice}]}{[\textit{coil}]}$$

Y a partir de la misma podemos determinar el la energía libre de formación de hélice como:



Esquema I. Esquema del equilibrio de dos estados *coil*-hélice.

Dicho modelo asume un cambio lineal de la energía libre del sistema (*ΔG)* con el agregado de TFE, el cual puede representarse por la siguiente ecuación:

$$\Delta G = \Delta G^{H2O} + m\ ([TFE]/[H_2O])$$

Donde ΔG^{H2O} es el cambio de energía libre de formación de α-hélice en agua, *m* es un parámetro que describe la magnitud del efecto del solvente en el cambio de la energía libre entre los dos estados [108] y [TFE]/[H_2O] es la relación molar entre TFE y H_2O.

En general, estas transiciones dan valores de ΔG^{H2O} positivos, ya que para péptidos se observa que la energía libre de la conformación hélice en buffer es mayor que la del estado *coil* (Esquema I).

En la figura I.7, se muestran en forma de gráficos de barras los datos de las energías libres obtenidas a partir del ajuste al modelo con su correspondiente análisis estadístico (Tabla I.2).

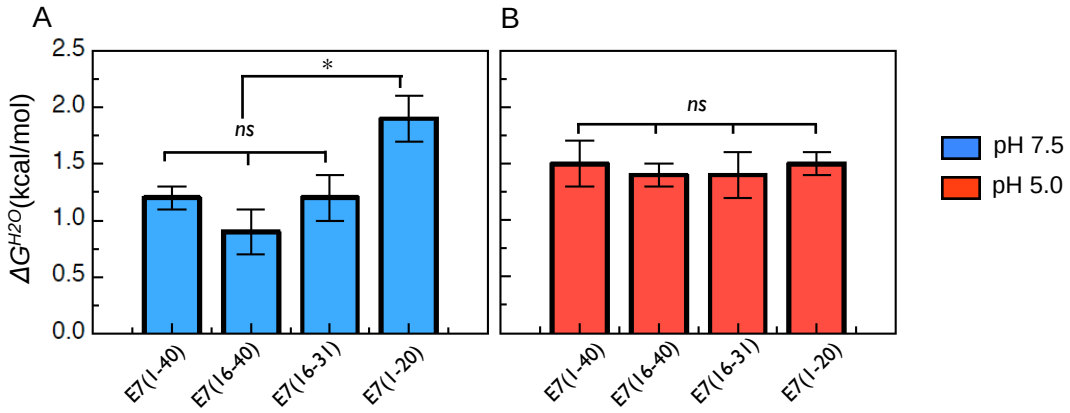


Figura I.7. Análisis de las energías libres de formación de α-hélice en agua a pH 7.5 y pH 5.0. Las barras representan los ΔG^{H2O} de formación de hélice en agua obtenidos a partir del ajuste de los datos de las titulaciones con TFE (Tabla I.2) a pH 7.5 (A) y a pH 5.0 (B). N=2. (ns, no significativo; * $p < 0.05$; ANOVA de una vía).

En primer lugar, al comparar los ΔG^{H2O} a pH 7.5 podemos observar que la hélice estabilizada en la región CR2 de E7N (observada para los fragmentos E7(16-40) y E7(16-31)) no presenta diferencias significativas entre su estado aislado o en el contexto del dominio completo (fragmento E7(1-40)) (Figura I.7-A). En contraste, la hélice estabilizada en la región CR1 (fragmento E7(1-20)) en estado aislado es significativamente menos estable que ambas hélices presentes en el dominio E7N (Figura I.7-A). Sin embargo, al comparar los ΔG^{H2O} a pH 5.0 no se observan diferencias significativas entre los distintos fragmentos (Figura I.7-B). Es interesante destacar que cuando se compararon los ΔG^{H2O} para un mismo fragmento a los dos valores de pH estudiados, no se observaron diferencias significativas para ninguno de los péptidos, incluido el fragmento E7(1-20). Estas diferencias, en conjunto, son difíciles de interpretar y requerirían un mayor número de repeticiones para obtener mayor poder estadístico. Proponemos que de existir dichas diferencias en la estabilidad, las mismas podrían sugerir que hélice localizada en E7(1-20) a pH neutro podría ser estabilizada mediante interacciones entre las regiones CR1 y CR2 transitorias. Esta hipótesis, si bien requiere un análisis más robusto, está apoyada por nuestras observaciones sobre los R_S extendidos medidos por AUC para los fragmentos E7(1-20) y E7(16-40) en comparación con E7N.

Por otro lado, a partir de los valores del elipticidad inicial y final obtenidos del ajuste, se calcularon los porcentajes de residuos en conformación α-hélice en H_2O y en TFE para los distintos fragmentos (Tabla I.2). Es importante destacar que debido a que el largo de los péptidos estudiados es variable, un mismo porcentaje de residuos en α-hélice para dos fragmentos de tamaño diferente corresponden a distinto número de residuos estabilizados en la hélice. (Por ejemplo, 20% de residuos puede equivaler a 20 residuos formando α-hélice si el péptido es de 100 residuos, o a 2 residuos en hélice si el péptido es 10 residuos).

Como es de esperar por las características desordenadas del espectro de CD en buffer (Figura I.4), el porcentaje de estructura α-hélice en H_2O es relativamente bajo para todos los fragmentos analizados (debajo del 6%) a ambos pHs (Tabla I.2). Dichos porcentajes aumentan a valores entre 12 y 20% a altas concentraciones de TFE para los distintos fragmentos (Tabla I. 2).

Tabla I.2. Parámetros de la estabilización α-hélice en E7N y los fragmentos

	pH 7.5				pH 5.0			
	ΔG^{H2O} (kcal/mol)	m (kcal/mol)	% α		ΔG^{H2O} (kcal/mol)	m (kcal/mol)	% α	
			H_2O	TFE			H_2O	TFE
E7(1-40)	1.2 ± 0.1	24 ± 3	4%	9%	1.5 ± 0.2	28 ± 3	6%	18%
E7(1-20)	1.9 ± 0.2	36 ± 4	6%	12%	1.5 ± 0.1	34 ± 2	6%	12%
E7(16-40)	0.9 ± 0.2	21 ± 4	2%	6%	1.4 ± 0.2	24 ± 4	6%	16%
E7(16-31)	1.2 ± 0.2	23 ± 4	6%	9%	1.4 ± 0.1	23 ± 2	6%	20%

El término %α representan el porcentaje de residuos en conformación α-hélice.

En la figura I.8 se muestran los porcentajes de residuos en α-hélice estabilizados en altas concentraciones de TFE a los dos pHs estudiados. En primer lugar, observamos nuevamente que los fragmentos aislados correspondientes a las regiones CR1 (E7(1-20)) y CR2 (E7(16-40) y E7(16-31)) pueden estabilizar estructura α-hélice en presencia de TFE (Figura I.8). Esto indica que en ambas regiones existen núcleos independientes para la formación de este tipo de estructura. En segundo lugar, al comparar el efecto del pH en la estabilización de estructura α-hélice en los fragmentos de E7N, observamos que el porcentaje de residuos en hélice para el fragmento del CR1 (E7(1-20)) no cambia con el pH. Sin embargo, todos los fragmentos que contienen o están incluidos en la región CR2, (E7(1-40), E7(16-40) y E7 (16-31)), duplican el porcentaje de residuos estabilizados al bajar el pH (Figura I.8). Una posible explicación a este fenómeno es que a valores de pH bajo, los carboxilatos de las cadenas ácidas de la región CR2 (Figura I.1), se neutralizan disminuyendo la repulsión electrostática y eliminando restricciones a la formación de estructura α-hélice, lo cual no ocurriría con la hélice del CR1 (ver más adelante).

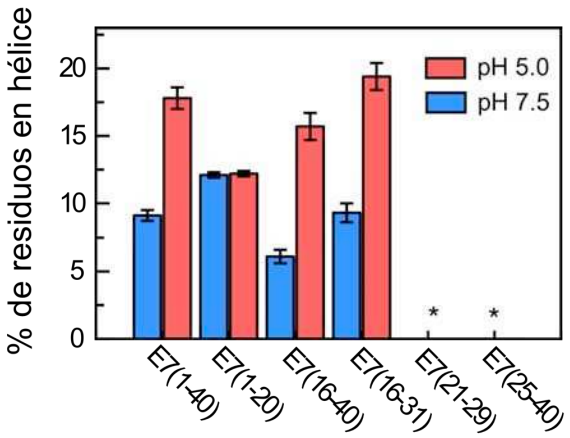
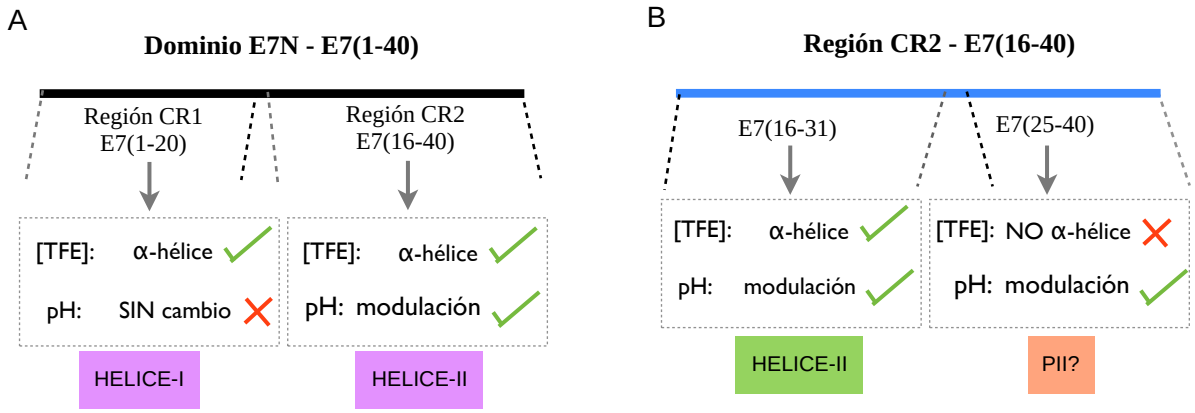


Figura I.8. Disección de elementos de α-hélices transitorias y locales dentro de E7N. Porcentaje máximo de α-hélice estabilizado en los distintos fragmentos de E7N medidos en buffer 10 mM Tris-HCl (pH 7.5) y 1 mM DTT, (barras azules) o 10 mM Formiato de sodio (pH 5.0) y 1 mM DTT, (barras rojas), ambos en la concentración máxima de TFE de la titulación. El asterisco indica que no hubo inducción de estructura de α-hélice.

En su conjunto, estos resultados demuestran que dentro del dominio E7N existen dos regiones que pueden estabilizar localmente y de forma independiente estructura secundaria α-hélice transitorias, que denominamos “Hélice-I” y “Hélice-II” por su ubicación respecto a las regiones conservadas CR1 y CR2 (Esquema II). La hélice-I está contenida dentro del fragmento E7(1-20) y el porcentaje de residuos estabilizados en dicha conformación no cambia al variar el pH entre 7.5 y 5.0 (Esquema II-A). Por otro lado, la hélice II esta contenida en el fragmento E7(16-40) y es modulada por cambios en el pH (Esquema II-B). Encontramos que la mínima región capaz de estabilizar a la hélice-II es el fragmento E7(16-31). Notablemente, estos resultados coinciden con las predicciones realizadas por AGADIR para el dominio (Figura I.5).



Esquema II. Estructuras secundarias transitorias estabilizadas en el dominio E7N (A) y análisis de estabilización dentro de la región CR2 (B).

Para profundizar en el efecto del pH sobre la estabilización de las hélices-I y -II, se realizó una titulación de pH (entre 3.0 y 8.0) para E7N y los distintos fragmentos. Se trabajó a una concentración de TFE cercana a la saturación para la mayoría de las titulaciones (30% TFE v/v) donde tanto la hélice-I como la hélice-II se encuentran estabilizadas (Figura I.6-B, línea punteada).

En primer lugar, podemos observar que el fragmento E7(1-20), en concordancia con lo expuesto anteriormente (Figura I.8), no presenta una transición estructural significativa con el pH (Figura I.9, columna 2), confirmando que la hélice-I no es modulada por el mismo. En contraste, los péptidos conteniendo la hélice-II; E7(1-40), E7(16-40) y E7(16-31), presentaron una transición estructural significativa con el pH (Figura I.9-A). Esto se observa por un aumento en la elipticidad molar a 195 nm hacia valores positivos, el corrimiento en el mínimo desde 198 hacia 208 nm y el concomitante aumento de la señal a 222 nm (Figura I.9-A, columna 1,3 y 4).

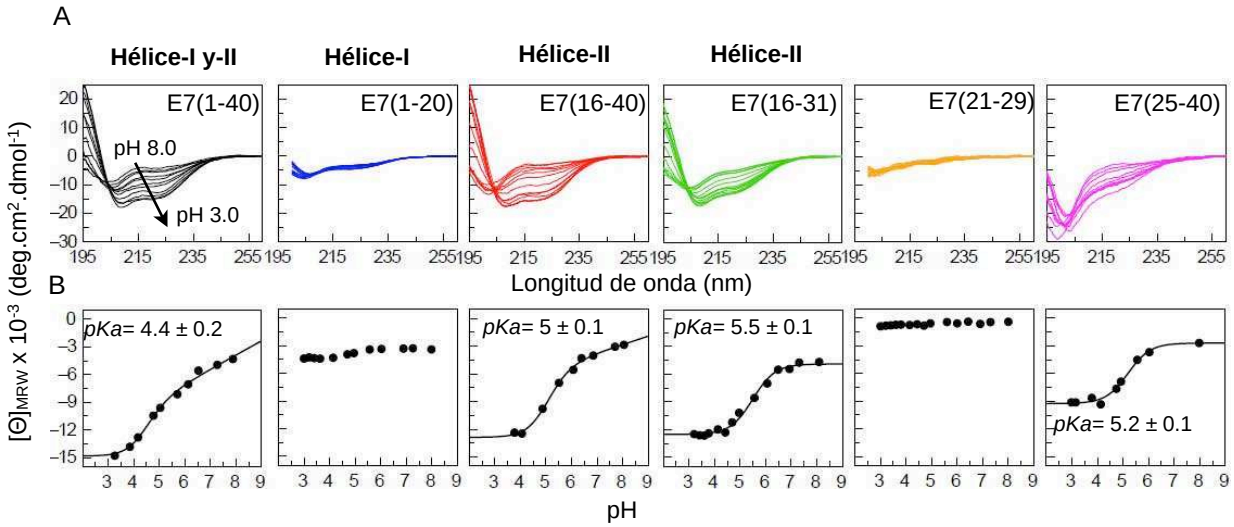


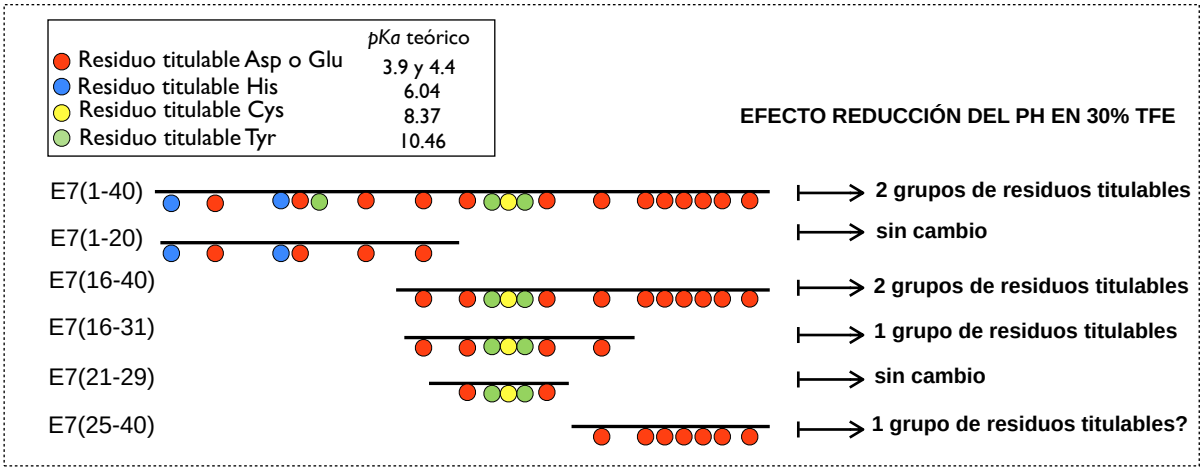
Figura I.9. Efecto del pH en la estabilización de estructura de los fragmentos de E7N. A-Espectros de CD en el UV lejano de los distintos péptidos de E7N a diferentes pHs. La flecha indica el sentido del cambio estructural al reducir el pH. B-Transición de pH al equilibrio seguidas por la variación de la elipticidad molar de la banda a 222. Todas las medidas fueron realizadas en buffer 10 mM citrato-fosfato en un rango entre pH 3 y pH 8.0 en 30% TFE y 1 mM DTT, a 20℃ y a una concentración de péptidos de 40 a 80 µM, luego de incubar las muestras por 1 hora a temperatura ambiente. Las curvas de pH fueron ajustadas a la ecuación de *Henderson-Hasselbalch* (Ver materiales y métodos). Dentro de cada gráfico se indican los valores de los pKa's estimados para cada fragmento.

Por otro lado, para obtener información cuantitativa sobre la transición inducida por pH se evaluó el cambio espectral de la banda a 222 nm en función del mismo. Se calcularon los *pKas* de las transiciones conformacionales inducidas por la protonación de los residuos titulables de cada fragmento, ajustando los datos a la ecuación de Henderson-Hasselbalch adaptada para un equilibrio de dos estados (Ver materiales y métodos). Como los fragmentos estudiados presentan más de un residuo titulable (Esquema III), los *pKas* obtenidos del ajuste corresponden a *pKas* globales o aparentes.

En la figura I.9-B se muestran los ajustes realizados para los fragmentos en donde se identificó transición estructural inducida por pH. Podemos observar que las transiciones observadas para E7N y de E7(16-40) presentan un *pKa* aparente de 4.4 ± 0.2 (Figura I.9-B, columna 1), y de 5.0 ± 0.1 (Figura I.9-B, columna 3), respectivamente. Una inspección detallada de ambas transiciones nos permiten detectar un ligero desvío del ajuste, sugiriendo la presencia de más de un grupo de residuos titulables. Por otra parte, el fragmento E7(16-31) presenta una transición monotónica con un *pKa* global de 5.5 ± 0.1 (Figura I.9-B, columna 4). La forma de dicha transición sumada a la presencia de un punto isodicroíco a 203 nm sugiere que todos los

residuos que están siendo titulados durante la transición estructural presentan un *pKa* similar (Figura I.9, columna 4). Finalmente, el fragmento E7(25-40) posee una transición con un *pKa* aparente de 5.2 ± 0.1. Sin embargo, el espectro de CD a pH 3.0 no presenta características de α-hélice, sugiriendo que la transición inducida por pH no involucra una transición del tipo *coil*-hélice (Figura I.9, columna 6).

Como se muestra en el esquema III, el dominio E7N contiene una elevada proporción de residuos potencialmente titulables en el rango de pH estudiados; residuos aspárticos (*pKa:* 3.9) y glutámicos (*pKa:* 4.4) (círculos rojos), histidinas (*pKa:* 6.04, círculos azules), tirosinas (*pKa:* 10.46, círculos verdes) y cisteínas (*pKa:* 8.37, círculo amarillo). Para identificar los residuos que se están protonando en el rango pH 3.0 y pH 8.0, se analizaron los desplazamientos químicos del Cβ entre pH 7.5 y pH 5.0 del dominio E7N por RMN (Figura apéndice A12). Se evaluó dicho núcleo porque es el núcleo asignado que se encuentra más cerca en la cadena a los grupos protonables y por lo tanto el que más resiente los cambios en el pH. Encontramos que los únicos residuos que cambian su entorno químico con el cambio de pH son los residuos ácidos y las histidinas (Figura apéndice A12). Ambas histidinas se encuentran localizadas dentro de la región CR1 que no es modulada por el pH, por lo tanto su protonación no esta implicada en la transición estructural estudiada. Sugerimos entonces que los residuos titulables que están reportando la transición estructural modulada por pH son los residuos ácidos presentes en la región CR2.



Esquema III. Detalle de los residuos con cadenas laterales potencialmente titulables dentro del dominio E7N.

Los valores de *pKa* calculados para cada fragmento (*pKa:* 4.4 ± 0.2, 5.0 ± 0.1, 5.5 ± 0.2 y 5.2 ± 0.1, para E7N, E7(16-40), E7(16-31) y E7(25-40), respectivamente) son ligeramente superiores a los valores teóricos reportados para los carboxilatos de los residuos aspárticos y glutámicos aislados (*pKa* 3.9 y 4.4, respectivamente). Proponemos que dichos *pKas* anómalos pueden deberse a un efecto del entorno químico de dichos residuos [109] o al efecto de la reducción de la constante dieléctrica del medio debido al agregado del TFE.

En resumen, sobre la base de los resultados obtenidos proponemos la existencia de dos grupos de residuos titulables la región CR2 de E7N cuya protonación modula equilibrios conformacionales. Uno de estos equilibrios estaría asociado a la estabilización de la hélice-II y el otro posiblemente asociado a la neutralización de los carboxilatos de la región ácida E7(25-40) (Esquema III).

I.1.6- Estabilización α-hélice por pH en el contexto de la proteína entera.

A fin de evaluar la influencia del dominio C terminal globular en la estabilización de α-hélice modulada por pH en E7N, se analizó comparativamente el efecto del cambio de pH en la estructura la proteína completa E7 y de una variante truncada E7(27-98). Se trabajó con dicha variante frente a la imposibilidad de trabajar con dominio E7C aislado, por su tendencia a formar oligómeros [67]. Sin embargo, dado que los residuos que comprenden la región bisagra entre el domino E7N y E7C son ricos en prolina (residuos 40-50) y por lo tanto presentan una conformación extendida, y que los residuos pertenecientes al fragmento E7(25-40) no estabilizan α-hélice incluso en 50% TFE (Figura I.6-A, columna 6), asumimos que los cambios observados sobre la variante E7(27-98) reportan exclusivamente cambios estructurales del dominio E7C.

Se midieron espectros de CD al UV lejano en buffer a pH 7.5 y pH 5.0 para la proteína entera, la variante truncada y el dominio E7N. En la figura I.10 se observa que las únicas dos especies que experimentan un cambio estructural entre pH 7.5 y pH 5.0 son el dominio E7N y la proteína entera (Figura I.10-A y C, líneas llenas). Esto se puede apreciar en ambos casos por la ganancia en valor absoluto de elipticidad de la banda a 222 nm y por el corrimiento del mínimo de CD ligeramente hacia la derecha (Figura I.10-A y B). Sin embargo, este cambio estructural con el pH no se observa para la especie truncada E7(27-98), indicando que el dominio responsable del cambio estructural inducido por pH es el dominio E7N.

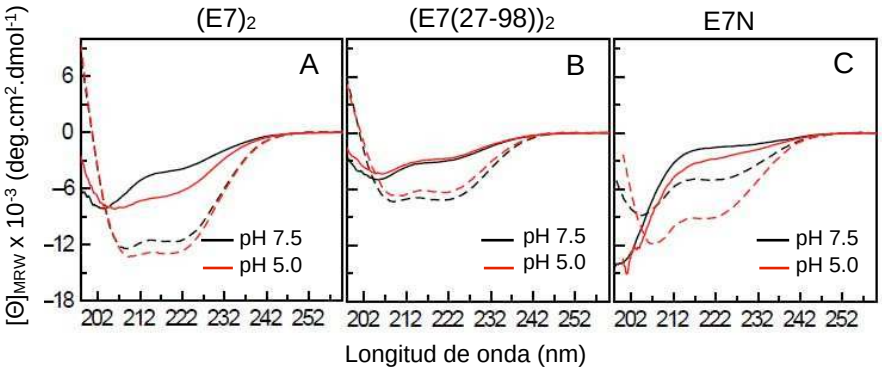


Figura I.10. Efecto de la presencia del dominio E7C en la estabilización por pH de la α-hélice en E7N. Espectros de CD en el UV lejano de $(E7)_2$ (A), la variante truncada (E7(27-98))2 (B) y el dominio E7N (C). Los espectros fueron medidos en buffer 10 mM Tris-HCl pH 7.5 (línea negra), en buffer 10 mM formiato de sodio pH 5.0 (línea roja), 10 mM Tris-HCl pH 7.5 y 1 mM DTT, con 50% TFE (línea negra punteada) y en buffer 10 mM formiato de sodio pH 5.0 y 1 mM DTT, con 50% TFE (línea roja punteada) a 20℃.

Por otra parte, el agregado de 50% TFE (v/v) estabiliza estructura α-hélice en la proteína entera, en el dominio E7N y en la variante truncada E7(27-98). Sin embargo, por lo expuesto anteriormente, interpretamos que la estabilización de estructura α-hélice observada por el agregado de TFE en E7(27-98) se debe a estabilización de estructura nativa del dominio E7C globular. Este supuesto se apoya en la diferencia en la magnitud de estabilización por TFE observada en la proteína entera en comparación con la proteína truncada.

En conjunto, estos resultados indican que E7N puede estabilizar conformación α-hélice inclusive en el contexto de la proteína entera, y por otra parte sugiere que E7N es el único dominio modulado por cambios de pH dentro de E7.

I.1.7- Estabilización de estructura secundaria en E7N por el detergente SDS.

El SDS es un surfactante que provee un entorno hidrofóbico que favorece la estabilización de distintas conformaciones [110]. A altas concentraciones forma micelas, y existen muchos ejemplos en literatura que describen que bajo dichas condiciones el SDS puede estabilizar estructuras α-hélice [110]. Por otro lado, también se ha demostrado que en concentraciones submicelares este detergente puede estabilizar estructuras tipo lámina-β [111]. Por estas razones, el mismo ha sido ampliamente utilizado para analizar tendencias conformacionales en péptidos *in vitro* [111]. En un trabajo de nuestro laboratorio se describió la capacidad del dominio E7N de adoptar estructura α-hélice en concentraciones supramicelares, así como la capacidad de adoptar conformación tipo lámina-β a concentraciones submicelares de SDS [64].

Para evaluar la capacidad de los distintos fragmentos de adoptar estructura tipo lámina-β, así como también para confirmar la existencia de las dos hélices locales identificadas en E7N (hélice-I y hélice-II), estudiamos el efecto del SDS sub- (1mM) y supra-micelar (25 mM) sobre la estructura de los mismos. Se trabajó a pH 3.0 para que todos los residuos ácidos se encuentren neutralizados. En dichas condiciones de buffer, la concentración micelar crítica del SDS es de 5 mM.

Al incubar los distintos péptidos con 1 mM SDS, encontramos que solo el fragmento correspondiente al dominio entero E7N adopta estructura lámina-β en dichas condiciones, lo cual se puede apreciar por la forma del espectro característico de estructura tipo lámina-β, con una banda positiva a 195 nm y de un mínimo alrededor de 218 nm (Figura I.11-A, columna 1). Esto demuestra que se requiere el domino E7N completo para estabilizar dicha estructura, sugiriendo la existencia de posibles interacciones entre las regiones CR1 y CR2.

Asimismo, a 1 mM SDS el péptido E7(1-20) estabiliza α-hélice mientras que los fragmentos E7(16-40), E7(16-31) y E7(25-40) presentan espectros de CD similares a los obtenidos en buffer (Figura I.11-A, columna 3 a 5). Esto sugiere que las hélices-I y -II responden de manera diferencial a la presencia de SDS, y por lo tanto presentan sensibilidad diferencial por distintos

entornos químicos. Por otro lado, a concentraciones supramicelares de SDS (25 mM), el dominio E7N, así como los fragmentos E7(1-20), E7(16-31) y E7(16-40) estabilizan conformaciones α-hélice (Figura I.11-B, columna 1 a 4). Por otro lado, al igual que lo observado para los experimentos en TFE, el fragmento E7(25-40), no estabiliza ningún tipo de estructura secundaria canónica inclusive a 25 mM SDS (Figura I.11-B, columna 5).

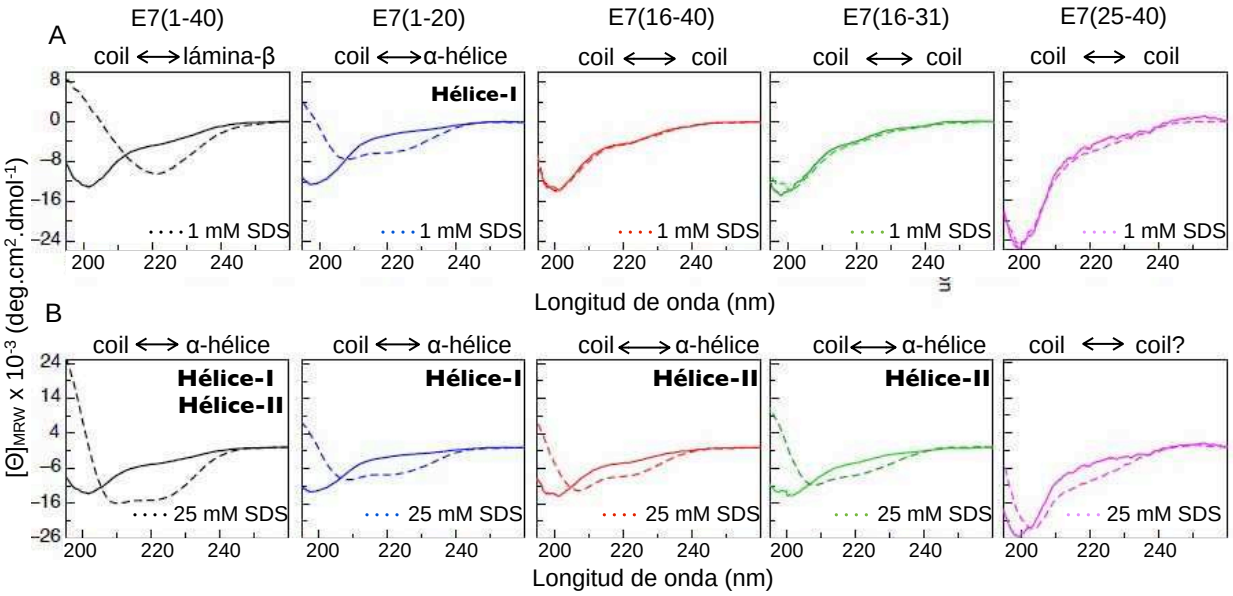


Figura I.11. Efecto del detergente aniónico SDS en la estructura secundaria de E7N y sus fragmentos. A Espectros de CD en el UV lejano de los distintos péptidos de E7N en buffer pH 3.0 con (línea punteada) y sin 1 mM SDS (línea llena). B-Espectros de CD en el UV lejano de los distintos péptidos de E7N en buffer pH 3.0 con (línea punteada) y sin 25 mM SDS (línea llena). Todas las medidas fueron realizadas en buffer 10 mM formiato de sodio pH 3 con o sin SDS, a 20℃. La concentración micelar crítica (CMC) calculada para el SDS en las condiciones estudiadas es de 5 mM.

Notablemente, el dominio E7N que contiene a las hélices-I y -II posee un espectro de CD con un mayor contenido de estructura α-hélice en comparación con los espectros de los fragmentos que contienen a las hélices individuales (Figura I.11-B, columna 1 a 4). Esta diferencia se puede apreciar mejor al analizar las señales crudas obtenidas para los espectros de CD del dominio E7N en comparación con la suma aritmética de los espectros individuales CR1 y CR2 determinados a una misma concentración (Figura I.12). Si los fragmentos CR1 y CR2 por separado estabilizan la misma cantidad de hélice que el dominio E7N a 25 mM SDS, no esperamos ver diferencias entre el espectro de CD de E7N y la suma de los espectros individuales correspondientes a las regiones CR1 y CR2. Sin embargo, en la figura I.12 se observa que el dominio entero estabiliza un mayor contenido de hélice que los fragmentos correspondientes a las regiones CR1 y CR2 por separado. Esto sugiere la existencia de posibles interacciones entre las hélices-I y II, que podrían impactar en una mayor estabilización de la estructura α-hélice dentro del dominio.

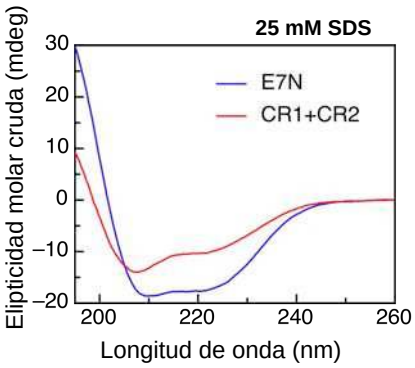


Figura I.12. Estabilización de estructura hélice en 25 mM SDS en el contexto del dominio E7N. Espectro de CD al UV-lejano del dominio E7N (línea azul) y de la suma aritmetica de los espectros crudos de la región CR1 y CR2 para los distintos péptidos medidos a una misma concentración de 30 µM en 10 mM formiato de sodio pH 3.0 con 25 mM SDS, a 20℃.

Estos resultados sugieren la existencia de interacciones de larga distancia dentro del dominio, y confirman la tendencia de los fragmentos que contienen a las hélices-I y -II a estabilizar dichas estructuras localmente.

I.1.8- Estructura de poliprolina tipo II en E7N.

Las hélices de poliprolina tipo II (PII), como se explicó en la introducción, son estructuras frecuentemente representadas en IDPs y en estados desplegados [39]. Por esta razón, la identificación de este tipo de estructura no es trivial, ya que la misma presenta características de cadenas desordenadas que son equívocamente identificadas con estados *random coil* [40]. Los espectros de CD al UV-lejano de péptidos con estructura PII presentan un mínimo a 198 nm y una banda positiva y diferencial a aproximadamente 218 nm. Se ha descripto previamente que bajas temperaturas promueven la estabilización de este tipo de estructuras, produciendo una disminución en la elipticidad a 198 nm y un aumento en la banda positiva a 218 nm [30].

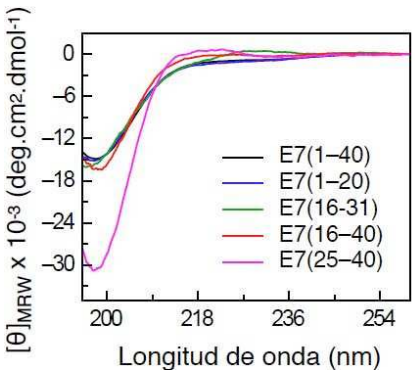


Figura I.13.Espectros de CD de los distintos fragmentos de E7N. Los espectros fueron medidos en buffer 10 mM Tris-HCl (pH7.5) y 1 mM DTT, a 5 ℃.

En la figura I.13 se muestran superpuestos los espectros de los distintos fragmentos de E7N medidos a pH 7.5 y a 5 ºC, donde se pueden apreciar las características distintivas de

E7(25-40) sobre el resto de los péptidos. Estas características indican la presencia de estructura PII.

Para analizar las regiones de E7N capaces de estabilizar este tipo de estructura, medimos espectros de CD de los distintos fragmentos a 5 ºC, 25 ºC y 75 ºC (Figura I.14). A 5 ºC, todos los péptidos presentan un mínimo a cercano a 198 nm y una banda de intensidad variable a 218 nm (Figura I.14-A). Esta señal a 218 nm desaparece al aumentar la temperatura a 75 ºC. Por otro lado, los valores de elipticidad a 198 nm aumentan, desplazándose levemente hacia longitudes de onda mayores. Para todos los fragmentos, los espectros diferencia muestran la aparición de la banda positiva a 218 nm y una disminución en la banda a 200 nm, sugiriendo la estabilización de estructura PII en todos los fragmentos (Figura I.14-B).

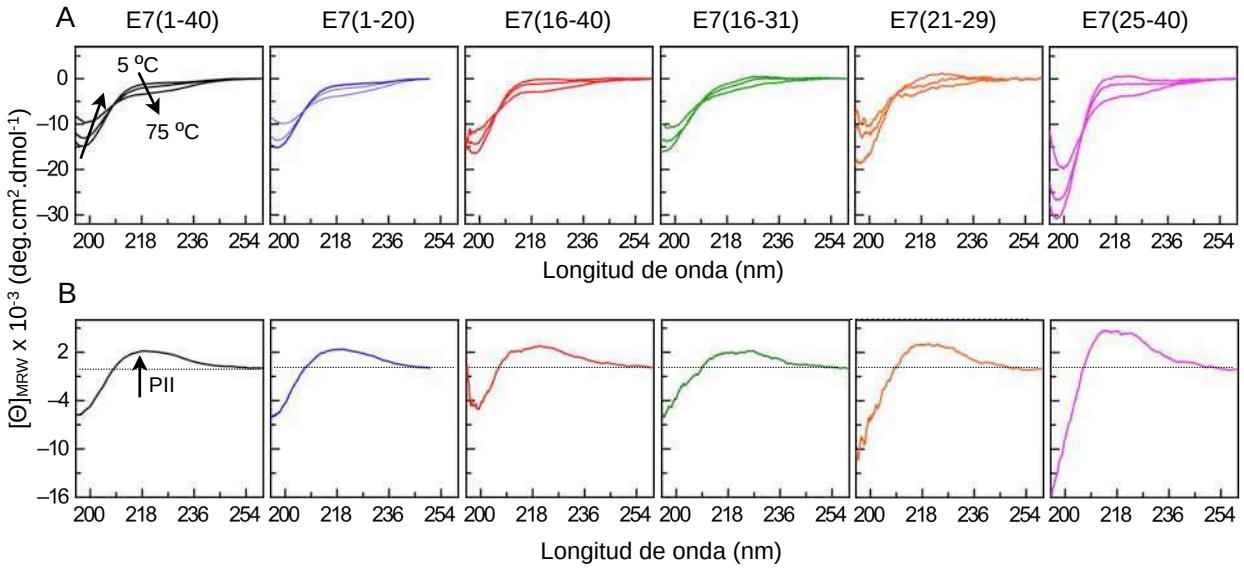


Figura I.14. Análisis de la tendencia de E7N a adoptar hélice de poliprolina tipo II por CD. A-Espectros de CD en el UV lejano de los distintos péptidos de E7N medidos a 5, 25 y 75 ºC. Las flechas indican el sentido de cambio conformacional con el aumento en la temperatura. B-Espectro diferencial entre 5 y 75 ºC, mostrando la estabilización de conformación PII a bajas temperaturas. La línea punteada indica el cero. La flecha muestra la inducción de la banda positiva a 218 nm, característica de estructura PII. Todos los espectros fueron medidos en buffer 10 mM Tris HCl pH 7.5 y 1 mM DTT, a una concentración de péptidos de 40 a 80 µM.

En concordancia con las observaciones anteriores, la magnitud de la inducción de PII resultó notablemente mayor en el fragmento E7(25-40), sugiriendo que dicha conformación se podría estar estabilizando por la repulsión electrostática entre los carboxilatos de los residuos ácidos presentes en el mismo (Figura I.1).

Para confirmar esta observación, se evaluó el impacto de la neutralización de los carboxilatos sobre la estabilización de estructura PII en los distintos fragmentos. Para ello analizamos el cambio en la magnitud de la banda a 218 nm obtenida a partir del espectro diferencia entre 5 y 75 ºC (Δ218 nm) a pH 7.5 y pH 5.0 (Figura I.15). Los únicos dos fragmentos que presentaron diferencias en el Δ218 nm a los dos pHs estudiados fueron los fragmentos que contienen la región ácida de E7N, el fragmento E7(25-40) y el fragmento E7(16-40) que lo incluye (Figura I. 15). Estos resultados apoyan la hipótesis de que la región de E7N con mayor tendencia a

estabilizar PII es la región ácida contenida en E7(25-40), y que la misma esta modulada por el pH, a través de la protonación de los carboxilatos. A pH 5.0 los carboxilatos de dicha región se encuentran mayormente protonados y por lo tanto se desfavorece la estabilización de la estructura PII.

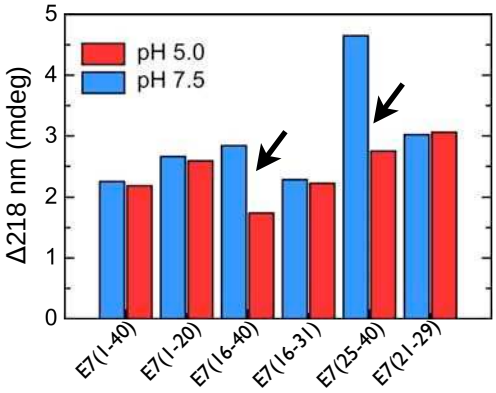


Figura I.15. Efecto del pH en la estabilización de estructura PII. Señal de la banda a 218 nm obtenida a partir del espectro diferencia entre 5 y 75 ºC para cada fragmento de E7N medidos en buffer 10 mM Tris-HCl (pH 7.5) (barras azules) o 10 mM Formiato de sodio pH 5.0 (barras rojas). Las flechas negras indican las posiciones donde se observa diferencias entre pHs.

Los resultados presentados hasta el momento indican que dentro del dominio E7N existen regiones con la capacidad de estabilizar localmente y en forma transitoria, elementos de estructura α-hélice y PII. En particular, identificamos dos regiones con tendencia a adoptar α-hélice a las cuales llamamos hélice-I y hélice-II. Ambas hélices fueron estabilizadas por distintos co-solventes (TFE, SDS) tanto en el dominio entero como en los sub-fragmentos que las contienen, demostrando que dichas hélices pueden formarse localmente y de forma independiente del resto del dominio. Sin embargo, a pesar de generarse localmente, los experimentos con SDS tanto a concentraciones sub- como supra-micelares sugieren la existencia de interacciones entre las regiones CR1 y CR2. Por otra parte, si bien todos los fragmentos en diferente medida presentan la capacidad de adoptar conformación PII, encontramos que la región que presenta mayor tendencia a adoptar dicha estructura es la región ácida E7(25-40). Sin embargo, a pesar de haber identificado condiciones donde el dominio E7N puede estabilizar de forma local elementos de estructura secundaria transitorios, las técnicas presentadas hasta el momento no nos permiten tener información a nivel residual de la extensión de los mismos. Por lo tanto, en la próxima sección buscaremos definir dichos elementos estructurales a partir de experimentos de RMN.

I.2.- DETERMINACIÓN DE LOS ELEMENTOS DE ESTRUCTURA SECUNDARIA DE E7N POR RESONANCIA MAGNÉTICA NUCLEAR.

Con el objetivo de obtener información a nivel atómico sobre los elementos estructurales locales identificados por CD al UV-lejano dentro del dominio E7N, se procedió a estudiar el mismo mediante RMN. El RMN es una técnica ampliamente utilizada para obtener información estructural de alta resolución en IDPs (ver Materiales y métodos).

Para proteínas en general, cuando una secuencia polipeptídica adopta una estructura canónica (α-hélice, lámina-β, etc) existen desviaciones sistemáticas para los desplazamientos químicos (CS, por sus siglas en inglés) de los núcleos del esqueleto peptídico respecto al valor esperado para ese núcleo si la secuencia estuviese en estado *random coil*. Por lo tanto, si se comparan las desviaciones que presenta un núcleo dado respecto a su valor esperado en un estado *random coil*, se puede obtener una indicación de la tendencia de una cierta región a poblar una estructura secundaria particular. Asimismo, la espectroscopía de RMN provee información de la tendencia de una cierta región de la proteína a poblar una conformación α-hélice o lámina-β, a través del análisis del patrón de picos cruzados observados en el espectro NOESY para los grupos amídicos de cada residuo, que es típico para cada tipo de estructura secundaria. Otra indicación que aporta el RMN acerca de la tendencia de una región de la proteína a estabilizar una estructura secundaria particular es brindada por los valores de las constantes de acoplamiento $^{1}JHC_{\alpha}$. Cuando el residuo esta en α-hélice esta contante presenta valores mayores que los observados para el mismo residuo en conformación *random coil*.

I.2.1- Asignación de las resonancias de RMN en E7N.

En primer lugar, se asignaron las resonancias de todos los ^{1}H y ^{13}C en abundancia natural para péptido E7(1-40) sin marcar mediante la combinación de los experimentos bidimensionales homonucleares y heteronucleares que se listan a continuación (ver materiales y métodos):

- **^{1}H-^{1}H NOESY**. (Espectroscopía de efecto nuclear Overhauser).
- **^{1}H-^{1}H TOCSY**. (Espectroscopía de correlación total).
- **^{1}H-^{13}C HMQC**. (Espectroscopía de correlación heteronuclear cuántica múltiple).
- **^{1}H-^{13}C HMQC-TOCSY**.

En las siguientes condiciones experimentales:

- **Condición I**. H_2O pH 7.5 (5% D_2O), 20 ºC. (Apéndice Tabla A1)
- **Condición II.** 50% v/v TFE-d_2/H_2O pH 7.5, 20 ºC. (Apéndice Tabla A2)
- **Condición III.** H_2O pH 5.0 (5% D_2O), 20 ºC. (Apéndice Tabla A3)
- **Condición IV.** TFE-d_2/H_2O 12% pH 5.0, 20 ºC. (Apéndice Tabla A4)

Observamos que E7N presentó baja dispersión de las resonancias de protones amida en el espectro monodimensional ^{1}H, con la mayoría de las resonancias en el rango de los 7.8 a 8.5 ppm. Esta es una característica esperada para una proteína IDP, debido al ambiente químico similar de los protones amida generado por la rápida interconversión de los confórmeros (datos no mostrados).

Se pudieron asignar todos los espectros a pH 7.5 y 50% TFE v/v. Sin embargo, a pH 5.0 en 50% TFE se observó agregación del péptido a lo largo de las determinaciones, llevándonos a trabajar a una concentración de 12% v/v TFE donde el péptido no mostró agregación y la estabilización de hélice todavía es significativa (Figura I.6, columna 1).

La figura I.16 muestra un ejemplo de una asignación realizada en solución acuosa a pH 5.0 (Condición III), utilizando experimentos ^{1}H-^{1}H y ^{1}H-^{13}C. Debido a la superposición de las señales en la región del protón amida, el análisis ^{1}H-^{13}C HMQC TOCSY fue esencial para asignar los distintos sistemas de *spin* (Figura I.16-A), mientras que la asignación secuencial de E7N se determinó analizando la conectividad NOE en los espectros ^{1}H-^{1}H NOESY (Figura I.16-B). Una observación interesante es que en los espectros de ^{1}H-^{13}C HMQC, las resonancias de los residuos se encuentran principalmente agrupadas por tipo de aminoácido, lo cual es característico de IDPs (Figura I.16-A).

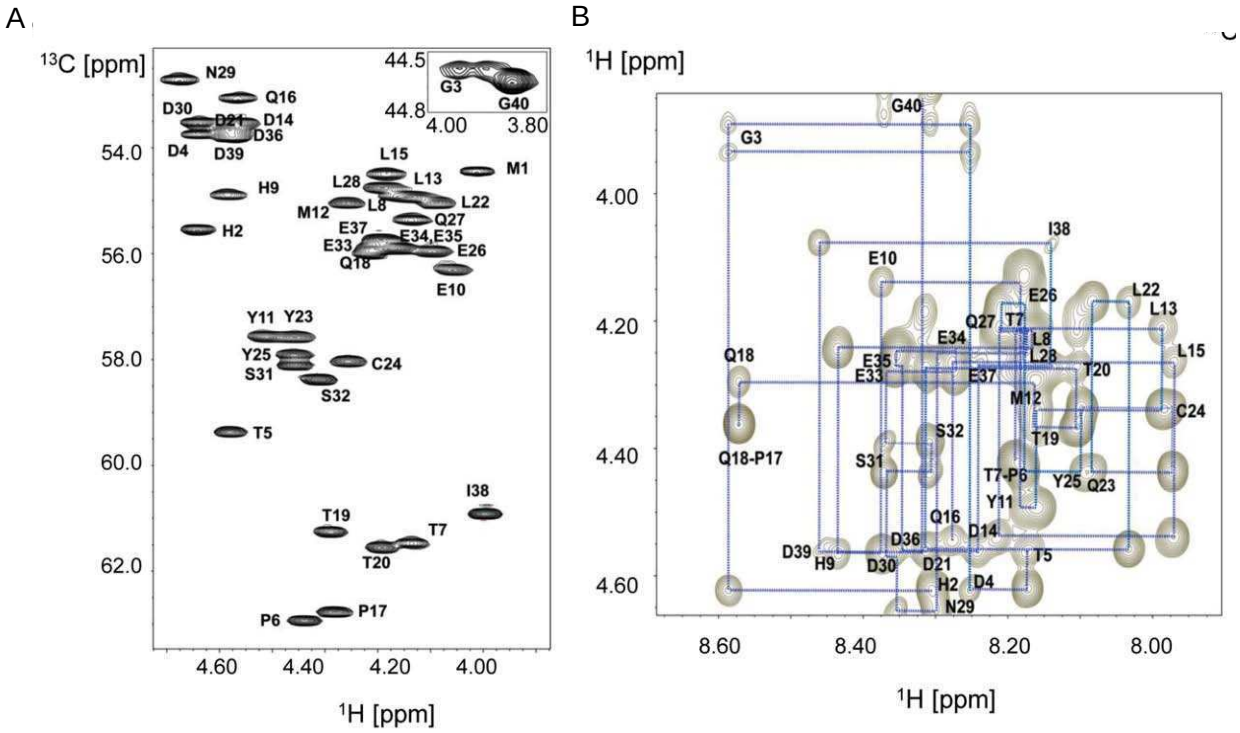
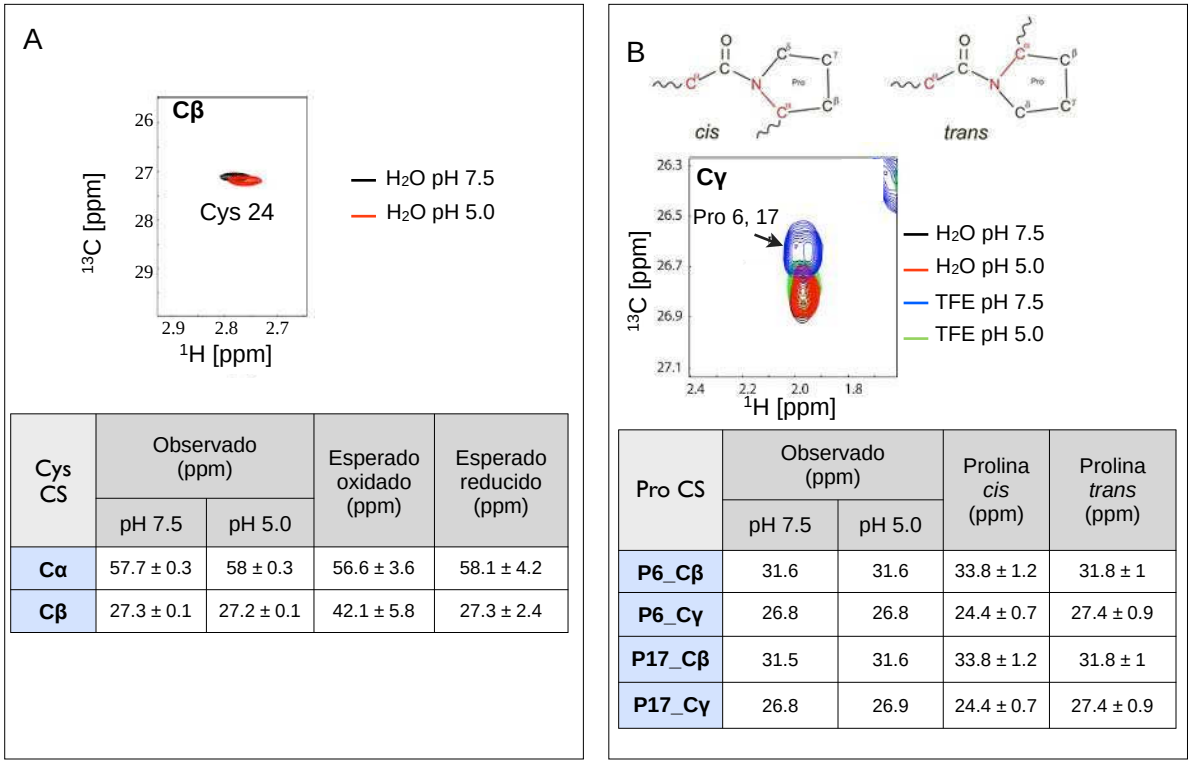


Figura I.16. Ejemplo de la asignación de los desplazamientos químicos de E7N. A-Espectro ^{1}H-^{13}C HMQC del péptido E7(1-40) en solución acuosa a pH 5.0. Las asignaciones de los *cross-peaks* $^{1}H_{\alpha}$-$^{13}C_{\alpha}$ se detallan en el interior de la figura. En el recuadro se muestran las asignaciones de las glicinas B-Región de la huella digital (backbone NH-$_{\alpha}$H) del espectro bidimensional ^{1}H-^{1}H NOESY indicando la conectividad secuencial NOE entre los αH. Por claridad, solo se detallan los *cross-peaks* NH-$_{\alpha}$H(i,i) y NH-$_{\alpha}$H(i+1,i). Todas las determinaciones de la figura fueron realizadas a una concentración de 3 mM de E7(1-40) en solución acuosa con 10 mM TCEP, a pH 5.0 y 20 ℃.

Por otra parte, analizando los desplazamientos químicos de los Cα y los Cβ confirmamos el estado reducido de la Cys24 de E7N en ambas condiciones estudiadas (pH 7.5 y pH 5.0) (Figura I.17-A). Además, es interesante destacar que en las condiciones experimentales en las cuales estudiamos al dominio E7N, sólo observamos un grupo de picos para cada prolina, P6 y P17, los cuales concuerdan con los parámetros tabulados para conformación *trans* (Figura I.17-B) [112]. Estos resultados contrastan con lo descripto para las dos prolinas presentes en la región bisagra entre los dominios E7N y E7C, las cuales presentan un equilibrio entre configuraciones *cis-trans* [66].



Cys CS	Observado (ppm)		Esperado oxidado (ppm)	Esperado reducido (ppm)
	pH 7.5	pH 5.0		
Cα	57.7 ± 0.3	58 ± 0.3	56.6 ± 3.6	58.1 ± 4.2
Cβ	27.3 ± 0.1	27.2 ± 0.1	42.1 ± 5.8	27.3 ± 2.4

Pro CS	Observado (ppm)		Prolina *cis* (ppm)	Prolina *trans* (ppm)
	pH 7.5	pH 5.0		
P6_Cβ	31.6	31.6	33.8 ± 1.2	31.8 ± 1
P6_Cγ	26.8	26.8	24.4 ± 0.7	27.4 ± 0.9
P17_Cβ	31.5	31.6	33.8 ± 1.2	31.8 ± 1
P17_Cγ	26.8	26.9	24.4 ± 0.7	27.4 ± 0.9

Figura I.17. Características generales del domino E7N por RMN en agua a pH 7.5 y pH 5.0. A-Porción de espectro ^{1}H-^{13}C HMQC mostrando los desplazamientos químicos obtenidos para el Cβ de la Cys 24 medida a pH 7.5 (línea negra) y a pH 5.0 (línea roja). En la tabla se indican los valores obtenidos de CS para el Cα y el Cβ, así como también los valores teóricos esperados para los distintos estados de oxidación de la Cys [166]. B-Porción de espectro ^{1}H-^{13}C HMQC mostrando los CS obtenidos para los Cγ de las Pro 6 y Pro 17. En el panel inferior se muestran los CS teóricos para las Pro en distinta conformación [112].

I.2.2- Identificación de α-hélices locales en E7N por análisis de CS.

Los desplazamientos químicos secundarios (desviaciones de los CS observados respecto a los valores tabulados de los residuos en *random coil*) son normalmente usados para evaluar tanto la localización como la estabilización relativa de elementos de estructura secundaria en proteínas. Sin embargo, la estructura secundaria en las IDPs es transitoria y por lo tanto la población total de una determinada estructura en el conjunto global de conformaciones de la proteína puede representar un porcentaje muy bajo. Dado que los CS representan una población media y ponderada de preferencias conformacionales locales, los CS secundarios

para IDPs son generalmente pequeños y por lo tanto difíciles de detectar [113]. Por otra parte, la falta de fiabilidad en los CS de las conformaciones *random coil*, pueden llevar a ambigüedades en el análisis [114]. Asimismo, los NOEs a corta distancia pueden utilizarse para corroborar las tendencias estructurales detectadas por el análisis de los CS. Sin embargo, en nuestro caso no fue posible debido a la superposición de señales en la región amida y alfa del espectro del protón. Por lo expuesto, con el objetivo de identificar a nivel de residuo la localización de las estructuras transitorias y locales identificadas en la sección anterior, se estudiaron los CS de E7N en distintas condiciones de pH y polaridad de solvente (condiciones I, II, III y IV). Analizamos las variaciones de los CS (ΔCS) entre pH 7.5 con y sin 50% TFE (condicion II - condición I) y entre pH 5.0 con y sin 12 % TFE (condición IV - condición III). Enfocamos nuestro análisis en los núcleos ^{1}Hα y ^{13}Cα, los cuales son los indicadores más robustos de estructura secundaria residual [115]. Las diferencias negativas en los CS del ^{1}Hα, así como diferencias positivas en el ^{13}Cα, indican estructura α-hélice [115].

En primer lugar, a pH 7.5 (condicion II - condición I) identificamos dos regiones dentro de E7N que presentan diferencias negativas en los ΔCS del ^{1}Hα, así como diferencias positivas en el ^{13}Cα (Figura I.18-A, recuadro violeta), indicando la presencia de dos regiones con tendencia a adoptar conformación α-hélice.

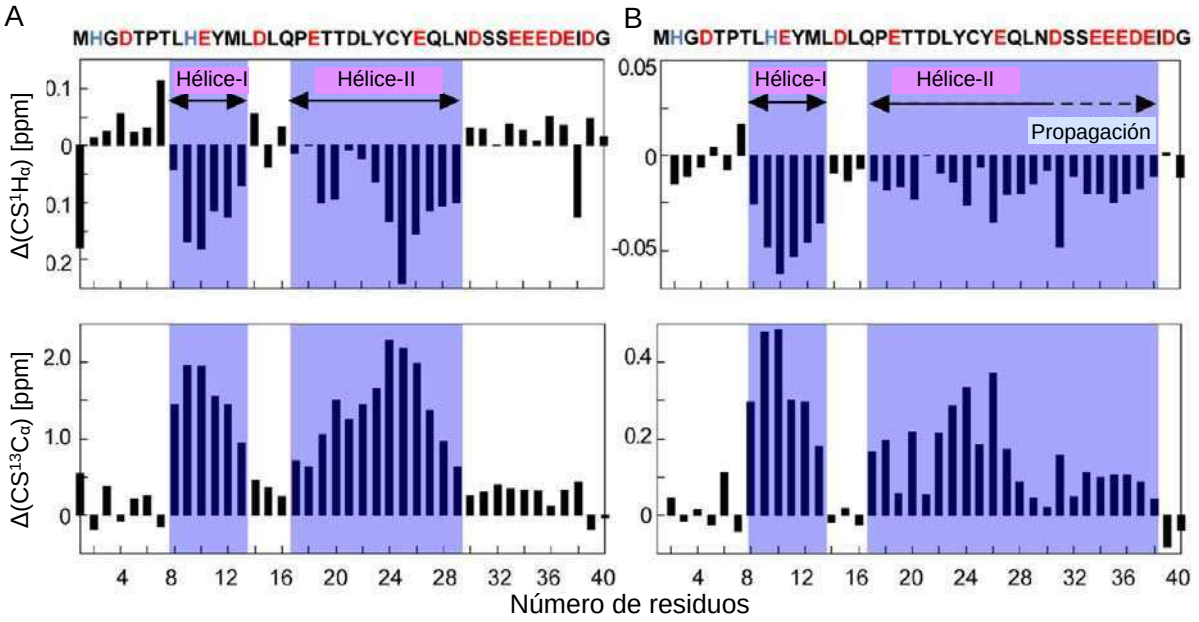


Figura I.18. Identificación a nivel de residuo de las α-hélices del dominio E7N por análisis de CS. ^{1}Hα y ^{13}Cα, corresponden a las diferencias de los desplazamientos químicos entre las soluciones agua y agua-TFE para 3mM E7(1-40) en 50% TFE-d_2 a pH 7.5 (A) y en 12% TFE-d_2 a pH 5.0 (B). Los contenidos de TFE agregados para la estabilización de hélice a los distintos pH fue diferente, y por lo tanto las magnitudes de los cambios de desplazamiento químico varían entre pH 7.5 y pH 5.0. Las regiones que estabilizan estructura α-hélice están indicadas en violeta.

La primera hélice comprende los residuos ^{8}LHEYML^{13} y la segunda esta conformada por los residuos ^{17}PETTDLYCYEQLN^{29}. Estos resultados confirman nuestras observaciones presentadas en la sección anterior sobre la existencia de las hélices-I y -II dentro de E7N,

además de permitirnos identificar la posición exacta de las mismas dentro del contexto del dominio entero.

Al analizar la composición de las hélices, podemos identificar 19 residuos involucrados en dichas estructuras (6 de la hélice-I y 13 de la hélice-II). Por otro lado, en los experimentos de CD en E7N (ver sección anterior) podemos observar un 9% de residuos estabilizados en esta condición (Tabla I.2). Este 9% de residuos puede ser interpretado como el 100% de moléculas de E7N formando una α-hélice de 4 residuos [115]. Si tenemos en cuenta el número de residuos que componen las hélices-I y II, podemos apreciar que este valor es cinco veces menor al número de residuos identificados por RMN, sugiriendo que en estas condiciones las dos α-hélices se encuentran parcialmente pobladas en E7N, representando solo al 20% de la población de moléculas de la misma. Este resultado es coherente con los valores bajos de CS secundarios observados, que son consistentes con el muestreo de un espacio conformacional amplio, poblando solo transitoriamente estructuras helicoidales.

Por otro lado, se estudió el efecto de la neutralización de los carboxilatos sobre la estabilización de estructura α-hélice en E7N a pH 5.0 (condición IV - condición III). Podemos nuevamente observar que se estabilizan dos hélices dentro del dominio. Sin embargo, mientras que la primera hélice no modifica su extensión ^{8}LHEYML^{13}, la hélice-II que comienza en la Pro 17 es propagada hasta el residuo I38 (Figura I.18-B). Estos resultados indican que la región ácidica de E7N (residuos 30-40) presenta la capacidad de adoptar estados conformacionales α-hélice cuando sus carboxilatos son neutralizados.

Esta propagación de la hélice-II permite explicar en parte la estabilización producida por el pH sobre la misma región observada en los experimentos de CD (Figura I.9). El hecho que el fragmento E7(16-31) pueda estabilizar hélice al neutralizar los carboxilatos, pero carezca mayormente de la región ácida, sugiere que la neutralización de los residuos ácidos de la región CR2 no solo favorecen la propagación de la hélice-II, sino que también aumentan la población relativa de esta estructura. Sin embargo, dado que se usaron concentraciones distintas de TFE a distinto pH (50% y 12%), los resultados de RMN no nos permiten sacar conclusiones sobre el efecto del mismo sobre la población relativa de residuos de E7N en conformación hélice.

I.2.3- Identificación de estructura PII en E7N por RMN.

Las regiones que estabilizan conformaciones tipo PII son difíciles de identificar por RMN principalmente porque la cadena carbonada no exhibe desviaciones características de los CS de los protones o los carbonos del esqueleto peptídico como las que se observan para otros tipos de estructura secundaria como α-hélice o lámina-β [116]. Sin embargo, los ^{1}JCαHα son indicadores a nivel residual de estructuras secundarias tanto PII como α-hélice. Aquellas

regiones que presenten CS típicos de α-hélice y diferencias en sus $^{1}JC\alpha H\alpha$ positivas, son indicativas de regiones α-hélices, mientras que aquellas regiones con diferencias en sus $^{1}JC\alpha H\alpha$ positivas pero que no presenten desviaciones de CS, son candidatas a poblar estructura PII.

Con el fin de identificar a nivel atómico la región dentro de E7N que es capaz de estabilizar hélice tipo PII, medimos las constantes de acoplamiento escalar $^{1}JC\alpha H\alpha$ obtenidas de espectros HMQC en condiciones que estabilicen (4 ℃ y H_2O pH 7.5) (Tabla apéndice A5) o que desestabilicen (20 ℃ y H_2O pH 5.0) estructura PII. Se determinaron $\Delta^{1}JC\alpha H\alpha$ como las diferencias entre las constantes de acoplamiento escalar medidas en cada condición y los valores tabulados para estados *random coil*.

En primer lugar, analizamos las constantes de acoplamiento escalar $^{1}JC\alpha H\alpha$ y los $\Delta^{1}JC\alpha H\alpha$ a pH 7.5 y a 4 ºC, condición en donde hay mayor estabilización de estructura PII [30]. A pesar de la superposición de picos, pudimos medir cuatro constantes de acoplamiento escalar en la región ácida del dominio E7N; E32, E33, D36 y I38 (Figura I.19, panel superior), las cuales presentan valores positivos respecto a los valores tabulados para conformaciones *random coil*. A pH 7.5 la región acídica de E7N presente valores de $\Delta^{1}JC\alpha H\alpha$ positivos, lo cual sumado a la ausencia de señales de conformación α-hélice analizadas por CS inclusive a concentraciones de TFE altas (Figura I.18), indica que esta región esta estabilizando estructura PII.

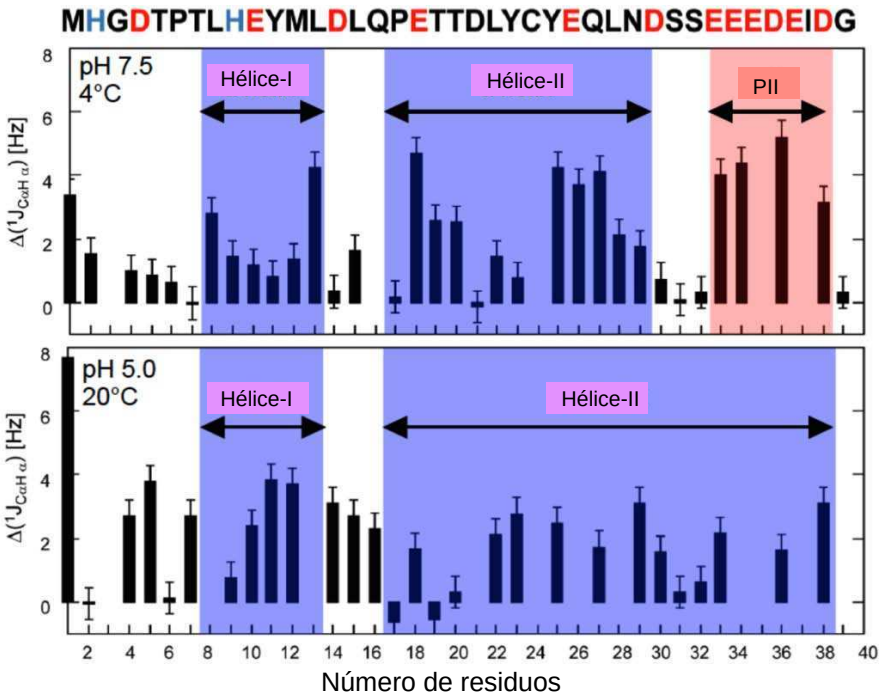


Figura I.19. Identificación a nivel residual de la región con tendencia a adoptar PII en E7N. Gráfico de la diferencia entre los valores de las constantes de acoplamiento escalar $^{1}JC_{\alpha}H_{\alpha}$ observados y tabulados para *random coil* a lo largo de la secuencia del dominio E7N, medidos a pH 7.5 a 4 ºC (panel superior) y soluciones acuosas a pH 5.0 a 20 ºC (panel inferior).

Por otro lado, las desviaciones positivas de Δ^1JCαHα coinciden con las regiones capaces de estabilizar las hélices I y II identificadas por CS a ambos pHs (Figura I.19, recuadro violeta). Sin embargo, en agua a pH 5.0 y 20 ºC, las constantes presentan desviaciones positivas desde H9 a Q16 para la hélice I y desde L22 a I 38 para la hélice II. Las pequeñas discrepancias entre las posiciones determinadas por CS y Δ^1JCαHα, pueden deberse al uso de valores *random coil* que no tienen en consideración los efectos de carga de la cadena y del entorno aminoacídico [117]. Estos resultados sugieren fuertemente que la región ácida del dominio E7N comprendida entre los residuos ^{33}EEEDEI^{38} adopta conformación PII a pH 7.5.

En conjunto, nuestros resultados de experimentos de CD y de RMN coinciden e indican que cuando los carboxilatos de la región ácida (33-38) contenida en el péptido E7(25-40) se encuentran desprotonadas a pH 7.5, la repulsión electrostatica de las cadenas laterales de los residuos ácidos estabiliza conformaciones extendidas tipo PII. Dicha conformación extendida genera restricciones conformacionales que llevan a que la hélice-II no pueda propagarse. Asimismo, a bajo pH cuando los carboxilatos se encuentran protonados, se pierde dicha estructura extendida, favoreciendo la propagación de la hélice-II.

DISCUSIÓN

Estudios previos de nuestro laboratorio demostraron que los primeros cuarenta residuos de la oncoproteína E7 de HPV-16 (E7N) constituyen un dominio intrínsecamente desordenado [64], el cual le confiere flexibilidad y propiedades hidrodinámicas particulares a la proteína entera E7. Se hipotetizó que dicha flexibilidad estructural podría explicar la capacidad de E7 de interactuar con los más de 100 blancos celulares y virales descriptos para la misma [62] [58] (Figura I.20).

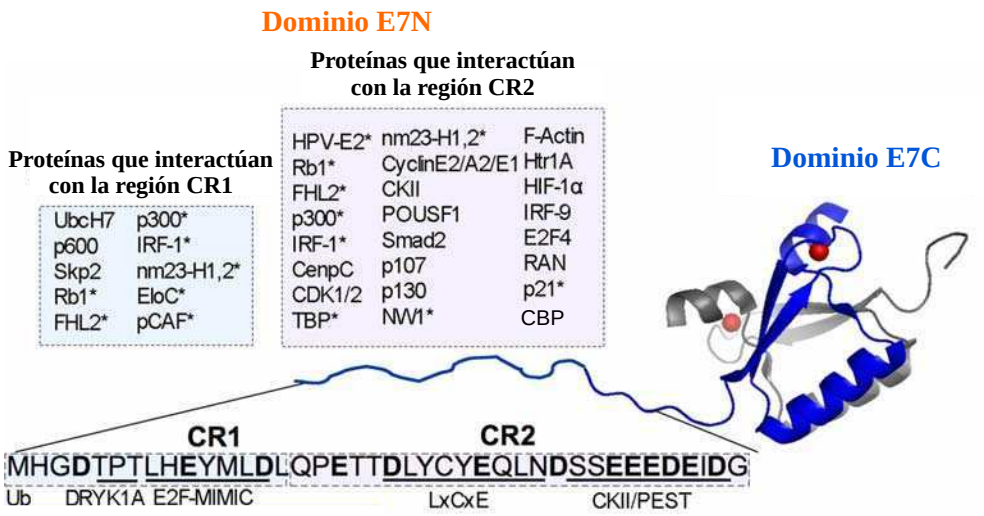


Figura I.20. Representación esquemática de la oncoproteína E7 de HPV-16. Los recuadros listan algunas de las proteínas blanco que interactúan con las regiones conservadas CR1 y CR2.

En el presente capítulo de esta tesis, se identificaron determinantes estructurales a nivel de residuos individuales que le otorgan plasticidad conformacional al dominio E7N de HPV-16. Definimos por CD y RMN tres elementos de estructura secundaria transitorios dentro del dominio E7N, los cuales se encuentran en equilibrio con un estado desordenado: (i) la hélice-I que abarca los residuos desde L8 hasta L13, dentro de la región CR1; (ii) la hélice-II que a pH 7.5 abarca los residuos P17 a N29 y a pH 5.0 es propagada hasta I38 y (iii) la hélice extendida PII estabilizada a pH 7.5, que comprende los residuos E33 hasta I38. Es importante tener presente que cuando nos refiramos a las hélices-I y -II a lo largo de esta discusión, nos referimos a las estructuras α-hélice transitorias que fueron identificadas por la estabilización con co-solventes como TFE y SDS, y por lo tanto son estructuras que se encuentran poco pobladas en las condiciones de buffer y temperaturas estudiadas inicialmente. De los experimentos de CD y RMN pudimos estimar que las mismas representan solo el 20% de la población de moléculas de E7N en alto TFE.

Efecto del pH sobre la hélice-I, la hélice-II y la región PII.

Como se mencionó anteriormente, una característica distintiva de las IDPs es la elevada proporción de residuos ácidos o básicos presentes en sus secuencias primarias, lo que provoca

que éstas presenten una carga neta distinta de cero a pH fisiológico [27]. Mientras que las proteínas globulares o con estructura definida tienden a perder estructura al ser sometidas a condiciones de pH extremo, las IDPs son capaces de ganar estructura [118]. Los cambios en la carga neta de una IDP, ya sea por modificaciones post-traduccionales, ligeras variaciones en el pH intracelular, o por la interacción con proteínas de carga opuesta, pueden inducir cambios estructurales que afecten su función biológica.

Trabajos previos de nuestro grupo de investigación describieron que E7 puede experimentar transiciones estructurales con cambios en el pH, y que el dominio E7N en particular presenta un aumento en el contenido de estructura α-hélice global a bajo pH [64]. Sin embargo, los detalles de las estructuras locales de dicho dominio, así como la interdependencia estructural entre los mismos no había sido explorada hasta el momento. En el presente capítulo demostramos la existencia de elementos de estructura secundaria transitoria y local en E7N, y definimos que el cambio global inducido por pH se debe a la modulación de dos de dichas estructuras transitorias. Observamos que existe una modulación diferencial de los distintos elementos estructurales transitorios descriptos en E7N.

En primer lugar, demostramos que la hélice estabilizada en la región CR1 (hélice-I) no es afectada por la protonación de las histidinas y los carboxilatos al disminuir el pH, lo cual es reflejado por la falta de cambios en la composición y/o el porcentaje de residuos en hélice en dichas condiciones y en presencia de TFE. Sin embargo, al analizar la estabilidad de la hélice-I aislada a pH 7.5, encontramos que los valores obtenidos de ΔG^{H2O} son significativamente mayores que los obtenidos para la hélice-II aislada o para la misma hélice-I en el contexto del dominio E7N (fragmento E7(1-40)). Podemos hipotetizar que dicha diferencia en la estabilidad de la hélice-I se debe a que en el contexto del dominio E7N la misma se podría encontrar estabilizada por interacciones transitorias entre la región CR1 y la región CR2. Evidencias apoyando la existencia de este tipo de interacciones fueron obtenidas por los experimentos de AUC (Tabla I.1), donde observamos que los R_S a pH 7.5 de los fragmentos E7(1-20) y E7(16-40) aislados son similares a los esperados para un fragmento mayormente desplegado en desnaturalizante, mientras que el dominio entero presenta un R_S intermedio entre una estructura extendida y su contraparte globular. A su vez, experimentos de estabilización de estructura con SDS también sugieren la existencia de dicho tipo de contactos transitorios (Figura I.11 y I.12).

Por otra parte, la hélice estabilizada en la región CR2 (hélice-II), presenta una modulación del contenido de hélice inducido por el cambio en el pH. Al neutralizar los carboxilatos de las cadenas laterales de los residuos ácidos a pH 5.0, encontramos que la hélice-II se propaga sobre la región ácida, comprendiendo los residuos P17 a D38. El hecho de que para E7N y para el fragmento E7(16-40) los valores de ΔG^{H2O} y los *m* obtenidos a ambos pHs sean similares sugiere que los residuos involucrados en la estructura PII no están involucrados en la

nucleación de una nueva hélice, lo cual es coherente con la hipótesis de una propagación de la hélice-II.

Por otra parte, los experimentos de CD realizados sobre el fragmento E7(25-40) muestran una disminución del contenido PII a pH 5.0. Proponemos que la neutralización de cargas relaja la conformación PII extendida favoreciendo la propagación de la α-hélice sobre esta región. Evidencias de este comportamiento fueron observadas en otros péptidos con tractos de poli-glutámicos, los cuales adoptan preferentemente estructura PII cuando sus carboxilatos se encuentran desprotonados [119].

Sin embargo, la propagación de la hélice-II permite explicar solo en parte la estabilización producida por el pH sobre la misma, observada en los experimentos de CD (Figura I.18 y I.9). El hecho que el fragmento E7(16-31) pueda estabilizar hélice al neutralizar los carboxilatos pero que carezca mayormente de la región ácida, sugiere que la neutralización de los residuos ácidos de la región CR2 no solo favorecen la propagación de la hélice, sino que también aumentan la población relativa de esta estructura. Sin embargo, se requieren más experimentos para demostrar esta posibilidad.

Es interesante destacar que encontramos que la hélice-I puede ser estabilizada en concentraciones submicelares de SDS en comparación con la hélice-II, la cual requiere de concentraciones supramicelares para estabilizarse. Dicha observación sugiere que ambas hélices presentan naturalezas distintas y por lo tanto pueden modularse diferencialmente en distintos entornos químicos.

Los resultados obtenidos demuestran la plasticidad conformacional que presentan las proteínas IDPs. Dentro de un mismo dominio, observamos una regulación diferencial de las estructuras secundarias locales mediante cambios en el pH, temperatura y polaridad del medio (SDS, TFE). Esta capacidad permitiría diversificar las funciones de una misma secuencia primaria de residuos, expandiendo el universo conformacional de una proteína frente a cambios en el entorno, permitiendo a una misma región de secuencia interactuar con más de un blanco distinto. Hipotetizamos que este puede ser el caso para el dominio E7N, el cual posee un gran número de blancos que interactúan con una misma región de la proteína (Figura I.20).

Participación de las prolinas en la estabilización de elementos de estructura transitoria.

Los residuos de prolina son uno de los dos aminoácidos naturales que no poseen ángulos típicos de torsión, ya que la forma de anillo restringe los ángulos Phi y Psi de la cadena carbonada. Las prolinas flanqueando estructuras canónicas α-hélice están asociadas a funciones estabilizadoras o desestabilizadoras de dicha estructura, dependiendo si se localizan hacia el extremo N- o C-terminal de la hélice, respectivamente [120]. En un trabajo reciente, se estudió el rol de las prolinas en la estabilización de regiones con estructura local α-hélice en

IDPs a través de un estudio computacional y de dinámica molecular, llegando a conclusiones similares [120, 121].

Al analizar la posición de las dos prolinas del dominio E7N (P6 y P17) respecto a las hélice-I y a la hélice-II, observamos que ambas hélices poseen una prolina hacia el N terminal. En particular, la hélice-I esta flanqueada por la P6 y la P17. Por otra parte, la hélice-II comienza en el residuo de P17 (Figura I.21).

En la figura I.21, se muestra un alineamiento de secuencias realizado sobre 139 secuencias pertenecientes a tipos virales del género alfa- y beta-papilomavirus, dicho alineamiento esta representado como un logo de secuencias [122] donde se muestra la conservación de residuos en el rango de 0 a 4.32 bits. Se puede apreciar que ambas P6 y P17 se encuentran conservadas en ambos géneros. Sin embargo, el grado de conservación de la P17 es notablemente superior. Esta observación sugiere que dicho residuo debe cumplir un rol estructural o funcional clave. Es interesante destacar que el fragmento E7(21-29) no es capaz de adoptar conformación α-hélice inclusive a concentraciones altas de TFE, lo que sugiere que el residuo que se requiere para la iniciación de la hélice no está contenido dentro de este péptido. Por lo tanto, proponemos que la P17 está funcionando como residuo iniciador de la hélice II, otorgando una mayor estabilidad a la misma.

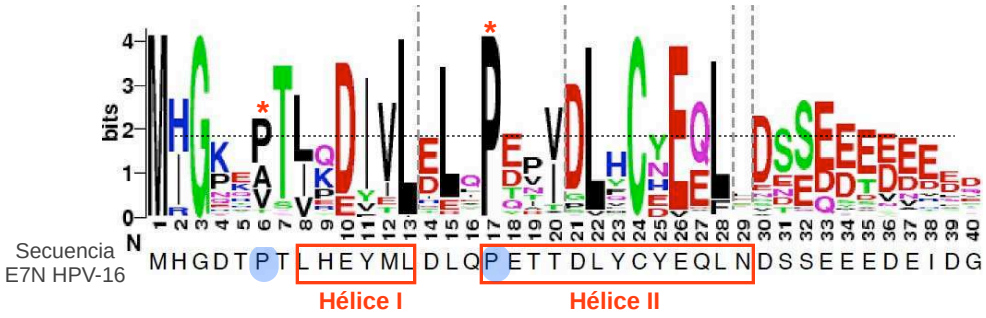


Figura I.21. Logo de secuencia de E7N. Se muestra un logo de secuencia del dominio E7N realizado a partir del alineamiento de 139 secuencias de HPVs pertenecientes a tipos virales del género alfa y beta. El recuadro indica la extensión de las hélices definidas por RMN. Los asteriscos rojos marcan las P6 y P17 del dominio. (Análisis realizado por la Dra. Chemes).

Por otra parte, en un trabajo del laboratorio (aún no publicado), se observó que una variante mutante de E7 en la cual se reemplaza la P17 por alanina, posee un mayor contenido de estructura hélice que su contraparte *wild type*. Esto sugiere que la P17 funcionaría no solo de iniciadora de la hélice-II, sino que también bloquearía la propagación de la hélice-I, funcionando como barrera estructural entre ambas hélices y posiblemente regulando sus niveles de helicidad. Un ejemplo de las prolinas como residuos que regulan los niveles de helicidad de elementos de estructura secundaria transitorios en IDPs fue descripto para la hélice de la región N terminal de p53 (13-27), la cual interactúa con MDM2 [123].

Los elementos de estructura definidos para E7N contienen a todos los motivos lineales de interacción descriptos para el dominio.

En general, los elementos de estructura secundaria transitoria presentes en las IDPs se encuentran involucrados en la interacción con diferentes blancos (revisado en [11]). Algunos ejemplos son las siguientes proteínas o dominios: p53 TAD, VP16 TAD, preS1 de HBV y 4EBP1, entre otros. En un trabajo reciente, se ha propuesto a dichos elementos de estructura secundaria transitoria o preformada como los "sitios activos" de las IDPs, en analogía con los sitios activos enzimáticos [11].

El dominio E7N interactúa con múltiples blancos celulares a través de sus regiones CR1 y CR2 (Figura I.22). Los motivos lineales presentes en patógenos "imitan" motivos celulares para intervenir dichas interacciones fisiológicas y secuestrar la maquinaria celular [55, 124]. El dominio E7N contiene 5 motivos lineales de interacción, y dada la cantidad de blancos identificados para estas regiones es intuitivo pensar que debe existir superposición de sitios de interacción dentro del dominio E7N.

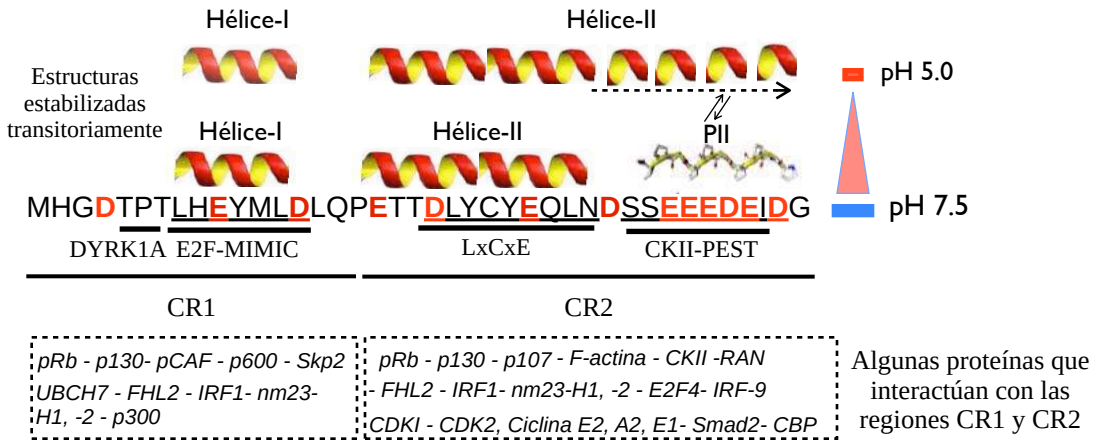


Figura I.22. Localización de los motivos lineales de interacción descriptos para E7N y los elementos de estructura secundaria transitorios. En las cajas debajo de cada región conservada se listan algunos de los blancos identificados para cada región.

Nuestros resultados nos permitieron localizar los elementos de estructura secundaria local estabilizados transitoriamente dentro de E7N respecto a los motivos lineales de interacción descriptos para el dominio. Todos los motivos lineales identificados hasta el momento, con la única excepción del motivo de fosforilación por la quinasa DYRK1A, se encuentran localizados dentro de la hélice-I, la hélice-II o la región PII (Figura I.22). El motivo E2F-MIMIC (residuos 8-14) coincide con la ubicación de la hélice-I, mientras que el motivo de unión a Rb LxCxE (residuos 21-29) se encuentra dentro de la hélice-II. Por último, los sitios de CKII (S31 y S32) y la región ácida PEST (residuos 30-40) se encuentran en la región que cambia entre PII y α-hélice según el estado de protonación de los carboxilatos (Figura I.22).

El motivo E2F-MIMIC es un motivo lineal presente en las familias de virus de tumores a ADN Papilomavirus y Adenovirus. Dicho motivo recibe ese nombre por imitar el motivo de unión a Rb del dominio de transactivación del factor de transcriptición eucariota E2F (E2F-TA). En la figura I.23-A se muestra un alineamiento de secuencia de ese motivo, y en forma comparativa la tendencia a hélice que posee esa región determinada por AGADIR. Se ha demostrado que dicha región, al formar complejo con Rb, adopta conformación hélice tanto en E2F-TA como en el motivo E2F-MIMIC de E1A (Figura I.23-B) [106, 125]. Nuestros resultados indican que la región correspondiente a este motivo lineal en E7 de HPV-16 (residuos 8-14), presenta la capacidad de estabilizar conformación α-hélice en presencia de co-solventes (hélice-I). Esto sugiere que la interacción del motivo E2F-MIMIC con RbAB podría implicar la selección de elementos estructurales preformados [126], en este caso la hélice-I, apoyando el modelo de selección conformacional para esta interacción en particular.

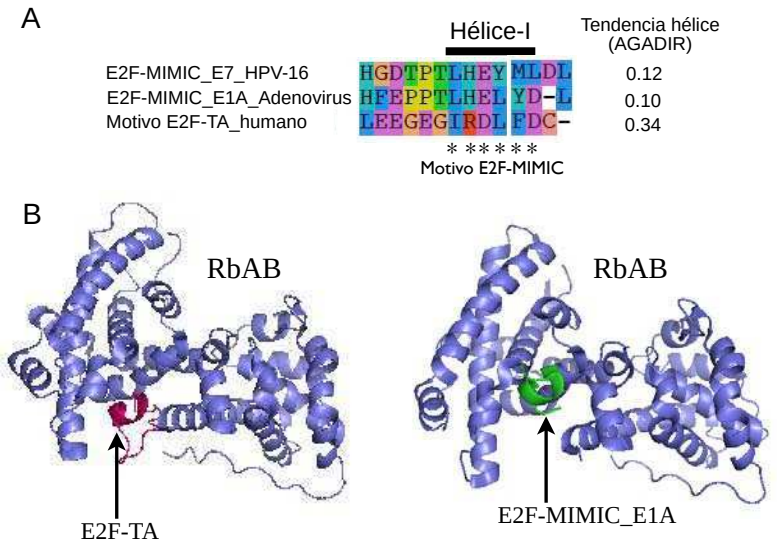


Figura I.23. Análisis comparativo del motivo E2F-MIMIC de E7 de HPV-16. A- Alineamiento de secuencias de los motivos E2F-MIMIC de E7 de HPV-16 y de E1A de Adenovirus, con el motivo celular de interacción con el dominio AB de la proteína retinoblastoma (RbAB) del factor de transcripción eucariota E2F1 (E2F-TA). B- Representación del complejo formado por RbAB y los motivos E2F-TA humano (izquierda, PDBID_1O9K) y E2F-MIMIC de E1A (derecha, PDBID_2R7G).

Nuestros resultados indican que el motivo LxCxE que fue descripto como el principal sitio de interacción con pRb, está contenido dentro de la región que puede adoptar conformación α-hélice en presencia de los co-solventes TFE o SDS. Sin embargo, dicho motivo adopta conformación tipo β-extendida al formar complejo con RbAB (Figura I.24-A) [127]. Las tendencias estructurales del motivo LxCxE podrían ser influidas en gran medida por el contexto de la secuencia vecina, por las condiciones del entorno, o por interacciones transitorias de largo alcance. La población de conformaciones que no corresponden al estado unido a Rb en esta región podría constituir un ejemplo del modelo de desplegado funcional [128]. Alternativamente, la conformación hélice podría mediar la unión a blancos adicionales en esta región. Recientemente se ha descripto que la proteína E7 forma un complejo ternario con las

proteínas celulares pRb y CBP, interactuando con ambas a través del motivo LxCxE de cada monómero de E7 (Figura I.24-B) [129], resaltado la multiplicidad funcional de esta región.

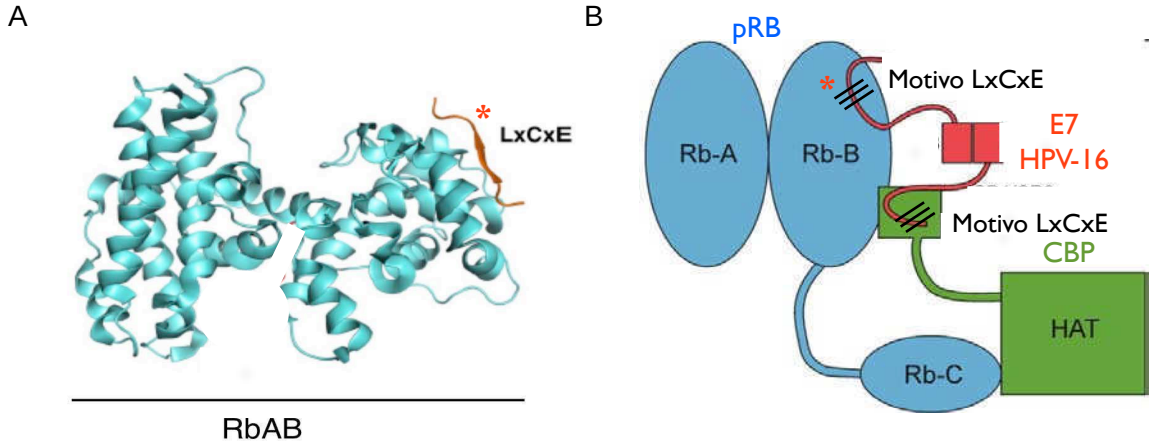


Figura I.24. Información estructural acerca de las interacciones mediadas por el motivo LxCxE. A-Estructura del dominio pocket de la oncoproteína RB (RbAB) en complejo con el péptido E7(21-29) correspondiente al motivo LxCxE de E7 (PDB ID: 1GUX) B-Modelo del complejo ternario propuesto para E7-pRb-CBP [129].

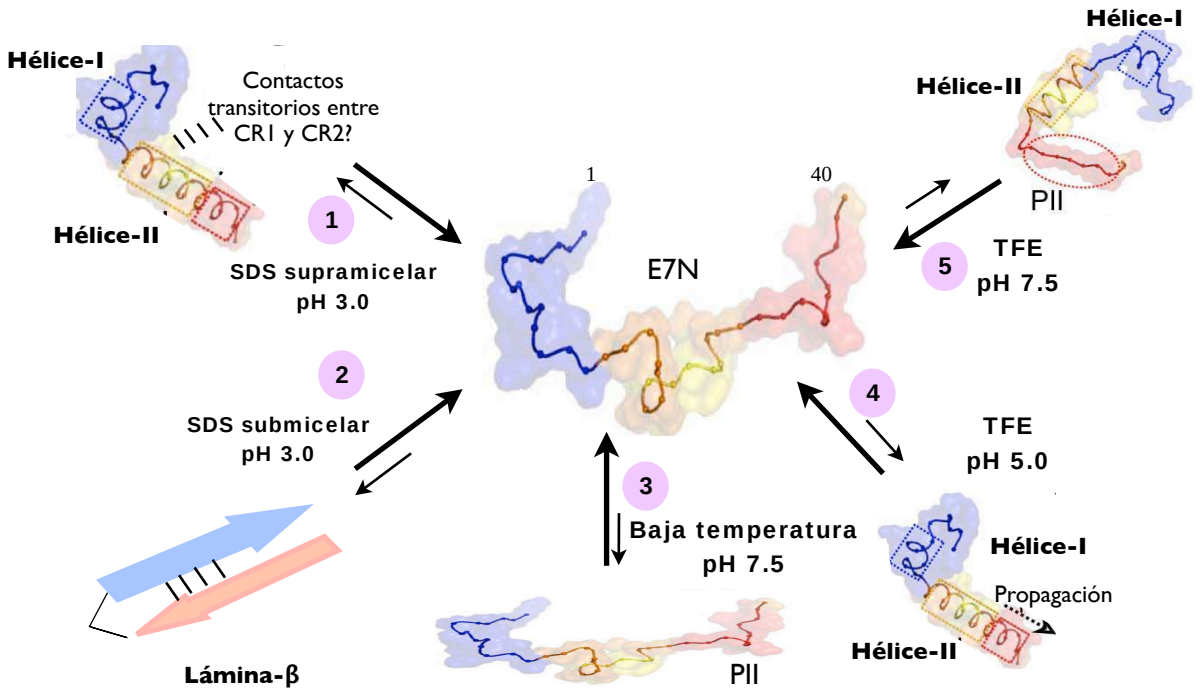
Por otro lado, los sitios CKII y PEST son contiguos en la mayoría de las proteínas E7 de HPV, muestran co-evolución, y están acoplados funcionalmente a la interacción con Rb [65, 130]. Los sitios PEST son pequeños motivos que etiquetan a las proteínas para su degradación, siendo dichas regiones ricas en prolinas (P), ácido glutámico (E), ácido aspartico (D), residuos de serina (S) y/o treonina (T) [41]. La fosforilación de S31 y S32 dentro de la región CKII-PEST incrementa la afinidad de unión a RbAB y el contenido PII de esta región [131]. Los resultados de este trabajo demuestran que la presencia de carga negativa desestabiliza la hélice-II y aumenta el contenido de PII en la región de CKII-PEST. Dicha conformación extendida en el entorno del motivo LxCxE podríá estar favoreciendo la accesibilidad del mismo permitiendo explicar el aumento en la afinidad por RbAB observada en dichas condiciones.

Por otra parte, el equilibrio α-hélice-PII en esta región también puede estar relacionado a la degradación de proteínas. Un trabajo de nuestro laboratorio se demostró que la fosforilación de un motivo CKII-PEST dentro de una región "bisagra" intrínsecamente desordenada de la proteína E2 del virus de papiloma bovino (BPV-1), conduce a un aumento de la degradación de proteínas *in vivo* [41]. Este ejemplo introdujo un nuevo concepto por el cual la fosforilación no actuaría como una "etiqueta" para el reconocimiento por quinasas o de la maquinaria de degradación, sino que desestabilizaría la estructura local dentro de esta región, produciendo una disminución en la estabilidad termodinámica local que repercute en la accesibilidad del sitio CKII-PEST a la proteasa [41]. Aunque esta hipótesis debe ser probada para nuestro sistema, el intercambio entre α-hélice y la estructura PII inducida por las variaciones de pH fisiológicos o fosforilación podría dificultar la propagación de la hélice-II en la región ácida a pH 7.5, promoviendo la accesibilidad del sitio PEST en forma de PII, con consecuencias directas sobre la tasa de degradación de E7.

Modelo de las posibles transiciones conformacionales del dominio IDP E7N

A continuación se muestra el modelo propuesto a partir de nuestros resultados, en donde se indican las regiones con estructura local que puede adoptar el dominio frente a distintos entornos. En resumen, en este capitulo demostramos que el dominio E7N existe en equilibrio entre sub-poblaciones con elementos de estructura secundaria local (α-hélice, PII, y posiblemente lámina-β) y con el estado "nativamente desplegado".

Modelo del equilibrio de posibles conformaciones experimentadas por el dominio intrínsecamente desordenado E7N



Definimos tres elementos de estructura secundaria local a los que denominamos helice-I, hélice-II y hélice PII. Demostramos que dichas estructuras pueden ser estabilizadas *in vitro* frente a diversas perturbaciones (cambios en la carga neta de la proteína, cambio en la polaridad del solvente, temperatura). La estabilización de estas estructuras *in vivo* podría estar dada por la interacción con diferentes blancos celulares, como por ejemplo la hélice-I al interactuar con pRb, o por la presencia de entornos hidrofóbicos como la proximidad a membranas. Asimismo, las observaciones mencionadas previamente sobre los R_S determinados mediante AUC, y el efecto del SDS sub- y supramicelar sugieren la existencia de interacciones de larga distancia presentes en el dominio E7N.

El ciclo de vida de los HPVs depende de la diferenciación de las células epiteliales basales a queratinocitos, lo que implica cambios sustanciales en el entorno físico-químico dentro de la célula, incluyendo parámetros tales como *crowding*, estado redox e incluso pH. Nuestros resultados sugieren que los estados conformacionales posibles de E7 están determinados por

equilibrios, y representan especies funcionales en la variedad de interacciones de las que participa esta proteína. Hemos demostrado que en distintas condiciones, E7N puede experimentar transiciones conformacionales entre estados α-hélice-*coil*-PII-lámina-β, y definimos las regiones involucradas, así como su estabilidad y su naturaleza transitoria. Estas transiciones podrían modular la exposición de sitios de modificacion post-traduccionales, sitios de degradación o sitios de interacción con distintas proteínas blanco. Esta plasticidad conformacional es la base para la interferencia de E7 con las vías de señalización claves requeridas para la transformación celular.

CAPITULO II

Caracterización bioquímica y biofísica de la fosfoproteína P de RSV.

CARACTERIZACIÓN BIOQUÍMICA Y BIOFÍSICA DE LA FOSFOPROTEÍNA P DE RSV.

La fosfoproteína P es la proteína sobre la cual se ensambla el complejo polimerasa de RSV, el cual es esencial para el ciclo de vida del virus. Esta característica hace de P de RSV un potencial blanco terapéutico. P es una proteína tetramérica, conformada por 241 aminoácidos, de la cual hasta el momento no se cuenta con estructura cristalográfica o por RMN de sus módulos o dominios. Estudios bioinformáticos y de proteólisis limitada sugieren que P presenta una organización modular compuesta por tres regiones [86]: el módulo N terminal (P_N; residuos 1-103), una región central (residuos 104-161) que incluye al dominio de tetramerización (P_Y; residuos 119-161), y el módulo C terminal (P_C; residuos 161-241). Estos estudios predicen que los módulos N y C terminales son intrínsecamente desordenados [86]. Sin embargo, no existen hasta el momento evidencias experimentales acerca de la naturaleza estructural de estas regiones. Por lo tanto, nos propusimos caracterizar bioquímica y biofísicamente los determinantes estructurales asociados a la flexibilidad conformacional de dicha proteína. Por lo expuesto anteriormente, planteamos la *hipótesis* de que P de RSV es una proteína con regiones intrínsecamente desordenadas. Por lo tanto, nos propusimos caracterizar la naturaleza IDP de esta proteína en solución mediante el abordaje de los siguientes objetivos:

- Identificar la/las regiones IDPs.
- Caracterizar el comportamiento estructural de cada región IDP aislada, así como también en el contexto de la proteína entera.
- Estudiar los distintos tipos de estructura secundaria que puedan estabilizar estas regiones.

Con el fin de desarrollar las herramientas necesarias para alcanzar los objetivos propuestos, se realizaron los clonados de las distintas construcciones, se expresaron y se purificaron los fragmentos recombinantes correspondientes a las distintas combinaciones de los módulos descriptos (Figura II.1) (ver Materiales y métodos). Brevemente, las especies P_N y P_C corresponden a los módulos individuales, N y C terminal, respectivamente. A su vez, las especies P_{NTET} y P_{TETC} corresponden a los módulos N terminal y C terminal, en el contexto del domino de tetramerización, respectivamente. La especie P_{C*} es un subfragmento resultante de proteólisis tríptica y corresponde a los últimos 44 residuos de P_C. Finalmente, también se produjo la proteína P entera y el dominio de tetramerización aislado (P_Y) (Figura II.1). Para el diseño de los distintos fragmentos se consideró la organización modular predicha para P, así como los datos experimentales que identificaron una zona central resistente a digestión por tripsina [86].

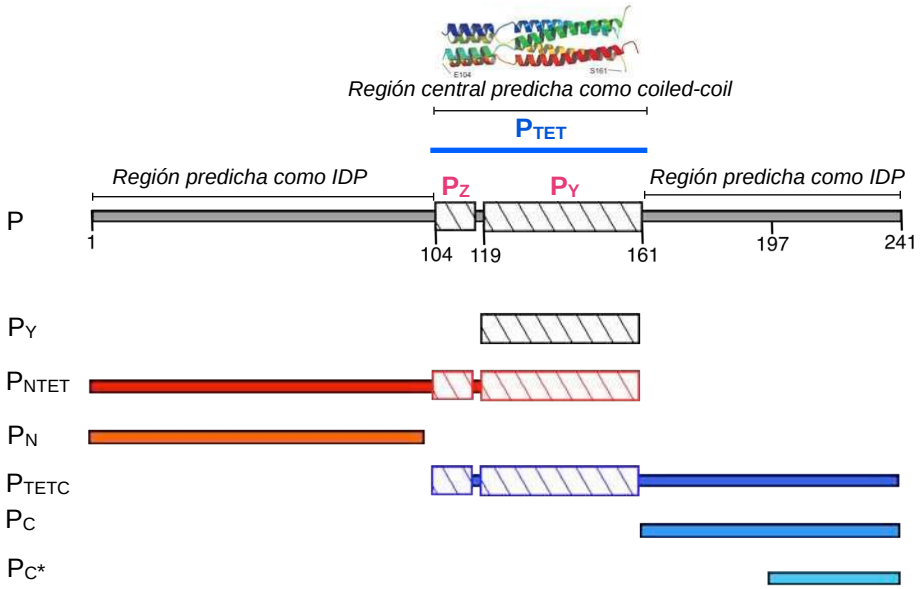


Figura II.1. Representación esquemática de proteínas y fragmentos recombinantes desarrollados en esta tesis. Las regiones predichas como desordenadas y *coiled-coil* están indicadas en la parte superior. En la parte superior de la figura se muestra el modelo tridimensional de la región central de P de RSV.

Notas de nomenclatura.

- Región central de P (P_{TET}): definimos que P_{TET} se encuentra conformado por los fragmentos P_Z y P_Y. Para esta definición nos basamos en el modelo propuesto para la región central de P [87, 88] (Figura II.1), y asumimos que la región 104-119 corresponde a una estructura tipo α-hélice. Por lo tanto, llamamos P_Z a la primer hélice de la región central comprendida entre los residuos 104 a 119 y P_Y al dominio de tetramerización propiamente dicho comprendido entre los residuos 119 a 161 (Figura II.1).
- Dominio N y C terminal de P: nos referimos a dominios N-terminal y C-terminal cuando hablamos de dichos módulos en el contexto tetramérico (P_{NTET} y P_{TETC}).

II.1.- ANÁLISIS HIDRODINÁMICO DE P Y LOS DISTINTOS FRAGMENTOS.

La cromatrografía de exclusión molecular en columna (SEC) es una herramienta sensible para determinar las propiedades hidrodinámicas de las proteínas, en particular, IDPs [31]. Por otra parte, utilizando la misma columna acoplada a un detector de dispersión de luz estática (SLS) se puede obtener el peso molecular de la proteína en estudio.

Para determinar el peso molecular nativo (MW_{SLS}) de cada una de las especies estudiadas, se realizaron medidas de dispersión de luz estática. En la figura II.2-A se muestra un ejemplo de esta determinación para el fragmento P_{NTET} (Figura II.2-A, línea roja). Las determinaciones de los otros fragmentos estudiados se detallan en la figura A10 del apéndice. Dado que P es una

proteína tetrámerica, determinamos el estado de oligomerización de cada fragmento estudiado. Para ello calculamos el cociente entre el peso nativo MW_{SLS} y el peso molecular teórico obtenido a partir de la secuencia primaria de la proteína utilizando el programa Expasy protParam (MW_{Expasy}). Todos los fragmentos que contienen el dominio de tetramerización resultaron ser tetrámeros en solución, mientras que los módulos aislados P_N y P_C son monómeros (Tabla II.1).

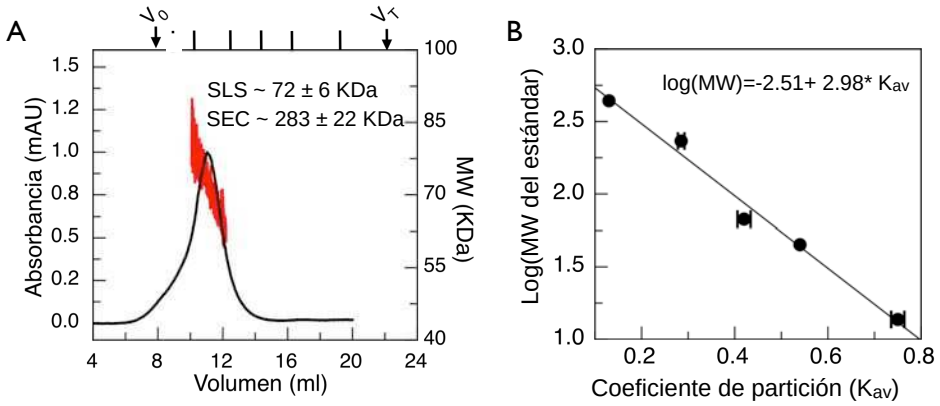


Figura II.2. Determinación del peso molecular de P_{NTET} por SEC y SLS. A-Cromatografía de exclusión molecular de P_{NTET} (línea negra). En la parte superior del gráfico se indican los volumenes de elución de los estándares de calibración. De derecha a izquierda: Ferritina (440 KDa), Catalasa (232 KDa), BSA (67 KDa), Ovoalbumina (43 KDa) y Ribonucleasa A (13.6 KDa). Las flechas indican las posiciones del V_0 y el volumen total (V_T) de la columna. En el eje de la izquierda se muestra la determinación de MW por SLS (línea roja). B-Curva de calibración utilizada para la determinación de los pesos moleculares aparentes por SEC. Todas las determinaciones por SEC fueron realizadas por triplicado (N=3), en buffer 20 mM Tris-HCl (pH7.5) y 200 mM NaCl a temperatura ambiente.

Por otra parte, para analizar el comportamiento hidrodinámico de P y sus fragmentos, los mismos fueron corridos en una columna de filtración molecular Superdex 200 analítica y se determinaron los pesos moleculares aparentes por SEC (MW_{SEC}) (Figura II.2-A, línea negra y Figura apéndice A10). Las determinaciones hidrodinámicas a partir de este tipo de corridas requieren que la columna sea calibrada previamente utilizando estándares de proteínas conocidas. En la figura II.2-B se muestra un ejemplo de la curva patrón generada a partir de la calibración y sobre la cual se determinaron los parámetros para poder calcular los pesos moleculares aparentes de los distintos fragmentos de P (Tabla II.1).

Tabla II.1.Determinación del peso molecular y del estado oligomérico de P y sus fragmentos.

Proteína	Peso molecular teórico (MW_{Expasy})	Peso molecular nativo (MW_{SLS})	Peso molecular aparente (MW_{SEC})	Estado oligomérico (MW_{SLS}/MW_{Expasy})
P	27.15 KDa	110.8 ± 1.9 KDa	421.4 ± 2.3 KDa	4.08 (Tetrámero)
P_Y	4.96 KDa	[132]	54.12 KDa	4 (Tetrámero)
P_{NTET}	18.5 KDa	72 ± 6 KDa	283 ± 22 KDa	3.9 (Tetrámero)
P_N	11.89 KDa	13 ± 1.4 KDa	32.7 ± 1.5 KDa	1.09 (Monómero)
P_{TETC}	15.89 KDa	62.7 ± 8.4 KDa	233.1 ± 1.5 KDa	3.94 (Tetramero)
P_C	9.26 KDa	9.3 ± 2.5 KDa	29.1 ± 1.8 KDa	1 (Monómero)

Una característica de las proteínas IDPs es su comportamiento hidrodinámico anómalo debido a que las mismas son extendidas en solución [16]. En general, presentan radios hidrodinámicos de Stokes (R_S) intermedios entre un estado completamente desplegado y su contraparte globular [31]. Existen evidencias del comportamiento extendido global en P de RSV [87, 132]. Por lo tanto, se procedió a determinar el/los módulos de P responsables de este comportamiento hidrodinámico anómalo, a fin de identificar y definir los módulos desordenados.

Para poder comparar las propiedades hidrodinámicas de fragmentos de diferente tamaño y estado de oligomerización, los valores de peso molecular aparente determinados por SEC (MW_{SEC}) fueron normalizados respecto al peso molecular nativo determinado por SLS (MW_{SLS}) (Tabla II.1), obteniendo el cociente MW_{SEC}/MW_{SLS}. Este cociente permite cuantificar el desvío del comportamiento hidrodinámico ideal esperado para una proteína globular (cociente(MW_{SEC}/MW_{SLS}) = 1) (Figura II.3, línea punteada). Todos los fragmentos presentaron cocientes (MW_{SEC}/MW_{SLS}) entre 2.5 y 3.9, consistentes con una conformación extendida en solución (Figura II.3). Los módulos monoméricos P_N y P_C presentaron cocientes (MW_{SEC}/MW_{SLS}) de 2.5 ± 0.1 y 3.1 ± 0.2, respectivamente (Figura II.3), sugiriendo la existencia de una conformación extendida y poco compacta de su cadena polipeptídica. Por otro lado, el dominio de tetramerización P_Y predicho como *coiled-coil* presenta un comportamiento anómalo con un cociente (MW_{SEC}/MW_{SLS}) de 2.7, debido a la naturaleza extendida y asimétrica esperada para las estructuras *coiled-coil* [87].

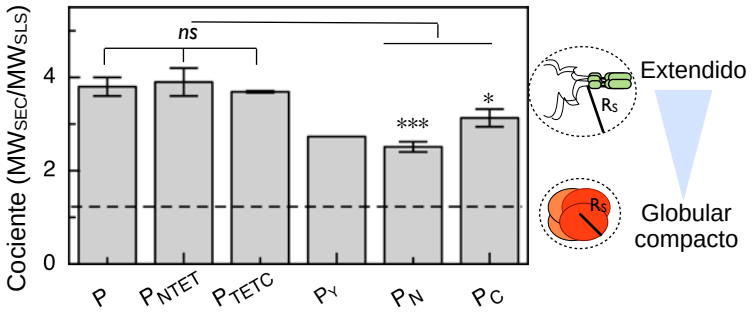


Figura II.3. Análisis del comportamiento hidrodinámico de los distintos módulos de P. Las barras representan el cociente entre el peso molecular determinado por SEC (MW_{SEC}) y el peso molecular determinado por SLS (MW_{SLS}). La línea punteada representa el cociente esperado para una proteína globular (MW_{SEC}/MW_{SLS} ~ 1). Todas las medidas fueron realizadas en buffer 20 mM Tris-HCl (pH7.5) y 200 mM NaCl a temperatura ambiente. Se representan la media y el desvio estándar de tres experimentos individuales (N=3). (*ns*, no significativo,* $p<0.05$, *** $p<0.001$, ANOVA de una vía).

Las especies P_{NTET}, P_{TETC} y P presentan desviaciones del cociente (MW_{SEC}/MW_{SLS}) significativamente mayores que P_N y P_C, con valores de 3.8 ± 0.3, 3.7 ± 0.1 y 3.9 ± 0.2, respectivamente. Estas desviaciones pueden ser explicadas por una combinación de conformaciones extendidas observadas en los módulos N y C terminal, y la presencia del dominio *coiled-coil*. Estos resultados demuestran que los tres módulos contribuyen al comportamiento hidrodinámico anómalo de P.

Por otra parte, calculamos los valores del radio hidrodinámico de Stokes (R_S) para los fragmentos monoméricos P_N y P_C y los comparamos con modelos de literatura [31]. Dichos modelos permiten determinar de manera teórica los R_S esperados para proteínas de un determinado peso molecular en conformación nativa y globular [Ecuación 1] y polipéptidos desplegados en agentes caotrópicos como Gdm.Cl [Ecuación 2], [103]:

[Ecuación 1] $\log(R_S) = -(0.204 \pm 0.023) + (0.357 \pm 0.005) \log(MW)$

[Ecuación 2] $\log(R_S) = -(0.723 \pm 0.033) + (0.543 \pm 0.007) \log(MW)$

donde los MW están expresados en Da y los R_S en Anstromgs.

Los radios hidrodinámicos de Stokes calculados para los módulos N y C terminal, P_N (R_S = 2.5 ± 0.2 nm) y P_C (R_S = 2.4 ± 0.2 nm), presentan R_S intermedios a los estimados para una proteína de 11.9 KDa o 9.3 KDa, respectivamente, en conformación globular (R_S = 1.8 o R_S = 1.6) y una proteína desplegada en Gdm.Cl (R_S = 3.1 o R_S = 2.7). Este resultado sugiere que a pesar de ser módulos extendidos en solución, P_N y P_C contienen elementos de estructura residual, apoyando la hipótesis de que estos módulos son intrínsecamente desordenados.

II.2.- LA FOSFOPROTEÍNA P DE RSV ES UNA IDP TIPO "MOLTEN GLOBULE".

Como se detalla en la introducción, una primera aproximación al análisis del contenido de desorden intrínseco de una proteína es evaluar la relación carga neta/hidrofobicidad media de su secuencia primaria [26]. Si bien no se puede conocer la estructura de una proteína a partir de su secuencia, el análisis de la misma puede predecir conformaciones no globulares [26]. A partir de dicha relación, se pudo establecer que aquellas proteínas o regiones de las mismas que presentan una relación carga neta alta/hidrofobicidad media baja, poseen una probabilidad alta de carecer de estructura secundaria canónica estable en condiciones fisiológicas [26]. La relación carga neta/hidrofobicidad media, está dada por la siguiente ecuación $\langle H \rangle_u = (\langle R \rangle + 1.109)/2.743$, donde $\langle H \rangle_u$ es la hidrofobicidad media y $\langle R \rangle$ es la carga neta media de una proteína a pH 7.0. Dicha recta divide al gráfico en dos regiones, por encima de esta se encuentran las secuencias que corresponden a las IDPs y por debajo de la misma, las proteínas estructuradas.

Como primera aproximación al análisis del contenido de desorden en la fosfoproteína P de RSV se evaluó la relación carga neta media/hidrofobicidad media [26] para la proteína completa P y los distintos fragmentos diseñados (Figura II.4). Como se puede observar, solo aquellas especies que contienen a los módulos N y C terminal de P presentan relaciones carga/ hidrofobicidad esperadas para proteínas IDPs, mientras que el dominio de tetramerización P_Y

coiled-coil se encuentra en el espacio conformacional predicho para proteínas estructuradas [26].

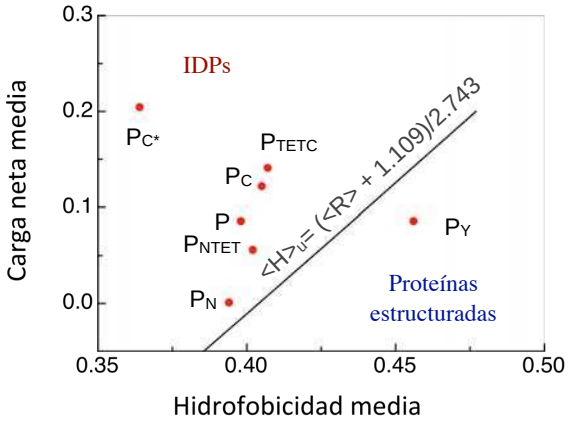


Figura II.4. Gráfico de carga neta *versus* hidrofobicidad de P y de los fragmentos utilizados.

Los resultados de la relación carga neta (pH 7.0) e hidrofobicidad media de P y sus fragmentos, predicen que los módulos N y C-terminales son IDPs y por lo tanto sugieren que los mismos carecen de estructura secundaria canónica estable en condiciones fisiológicas.

II.2.1- Análisis del contenido de estructura secundaria de P y de sus distintos módulos.

Para poder evaluar la contribución de cada módulo de P a la estructura secundaria global de la misma, se midieron en primer lugar los espectros de CD al UV-lejano (195 y 260 nm) de todos los fragmentos de P a 25 °C.

En la figura II.5-A se muestra el espectro de CD de P, el cual presenta contribuciones de estructura tipo α-hélice con el aporte de componentes de desorden intrínseco. Esto se puede explicar por la presencia dos mínimos definidos a 206 y 220 nm, y una relación entre la intensidad de sus bandas a 220/206 de 0.76 (Figura II.5-A). El aporte principal de estructura α-hélice al espectro global de P es dado por el domino de tetramerización P_Y el cual presenta un espectro característico de α-hélice, con dos mínimos de intensidad similar ubicados a 208 y 220 nm, y una banda positiva a 195 nm (Figura II.5-B) [87, 88].

Luego, evaluamos la contribución de los dominios N y C-terminal en la proteína entera. Observamos que el fragmento tetramérico P_{NTET} (Figura II.5-C) presenta un espectro de CD con características similares al observado para P, lo cual sugiere que a pesar de presentar un 30% menos de residuos (165 residuos/241 residuos totales en P), la proporción relativa entre desorden y orden se mantiene similar a la observada para la proteína completa.

Al evaluar el fragmento P_{TETC} observamos que el mismo presenta un espectro de CD con dos mínimos de 208 y 220 nm con una relación entre las bandas a 220/208 nm de 0.85, lo cual

sugiere un mayor contenido de estructura tipo α-hélice en comparación con P y P_{NTET} (Figura II. 5-D).

Estos resultados muestran que el dominio C terminal aporta un mayor contenido de estructura secundaria de tipo α-hélice a la estructura global de P, en comparación con el domino N terminal, sugiriendo que el módulo C terminal aislado contiene globalmente un mayor contenido de estructura secundaria que el módulo N terminal.

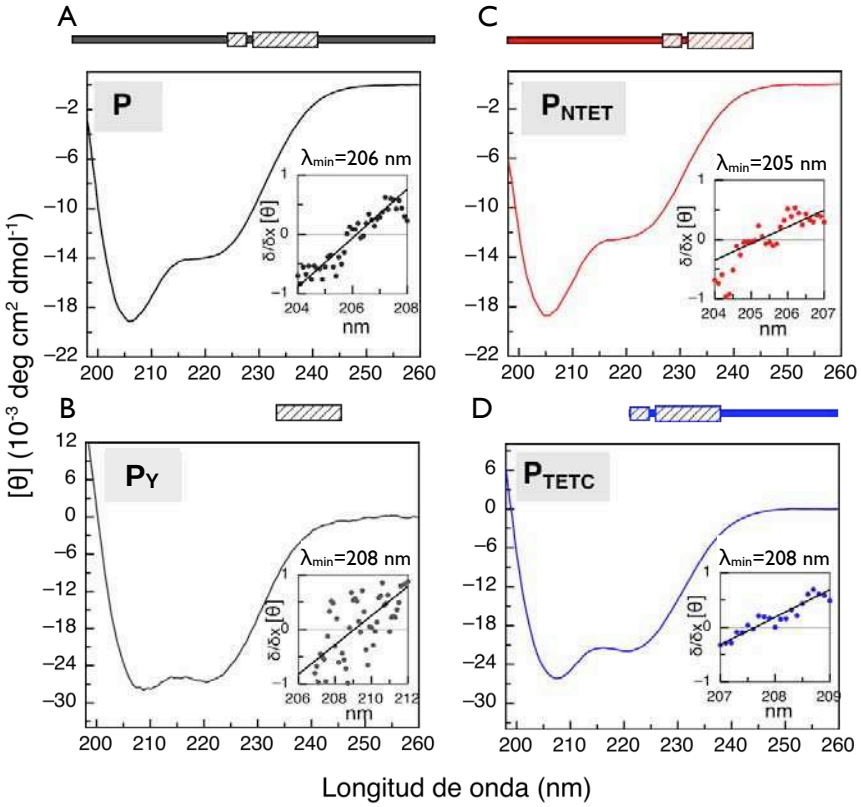


Figura II.5. Análisis del contenido de estructura secundaria por CD al UV lejano. Espectros CD de 10 µM de P (A) y los diferentes módulos tetraméricos P_Y (B), P_{NTET} (C) y P_{TETC} (D). Recuadro: determinación del mínimo de CD. Primera derivada de los espectros de CD. La posición del mínimo corresponde al punto donde la recta es igual a cero. Todos los espectros fueron medidos en buffer 20 mM Tris.Cl pH 7.5 y 50 mM NaCl, a 25 ºC.

Para confirmar esta observación, se analizaron los espectros de CD para P_N, P_C y para el sub-fragmento de C P_{C*}, en las mismas condiciones experimentales. Globalmente, las tres especies presentan espectros de CD típicos de proteínas desordenadas, con mínimos alrededor de 200 nm y sin componentes de estructura secundaria aparente (Figura II.6-A, B y C). Sin embargo, para el módulo P_C, la posición del primer mínimo a 202 nm y la presencia de un segundo mínimo a 220 nm indican la contribución de estructura α-hélice residual dentro de esta especie (Figura II.6-B), confirmando lo observado en el dominio C terminal de P. Dicho componente estructural no se observa en el sub-fragmento P_{C*} correspondiente a los últimos 44 residuos de P_C (Figura II.6-C), sugiriendo que la estructura estabilizada podría tratarse de una hélice local centrada en los primeros residuos de P_C, o de una hélice que requiere toda la información de secuencia del dominio C terminal para su formación.

Para descartar el efecto del buffer, fuerza iónica, y concentración en la estructura secundaria de P y sus fragmentos, se realizaron medidas en buffer fosfato, 0.02 a 0.3 M NaCl y en concentraciones entre 5 y 50 µM para las proteínas tetraméricas y entre 8 y 100 µM para los fragmentos monoméricos, no observando ninguna diferencia en el espectro de CD.

De este primer análisis concluimos que los aportes de desorden a la estructura global de P están dados por los módulos, P_N y P_C, los cuales en estado libre presentan espectros de CD característicos de IDPs. Sin embargo, el módulo C terminal presenta un mayor contenido de elementos de estructura secundaria tipo α-hélice en comparación con el módulo N terminal.

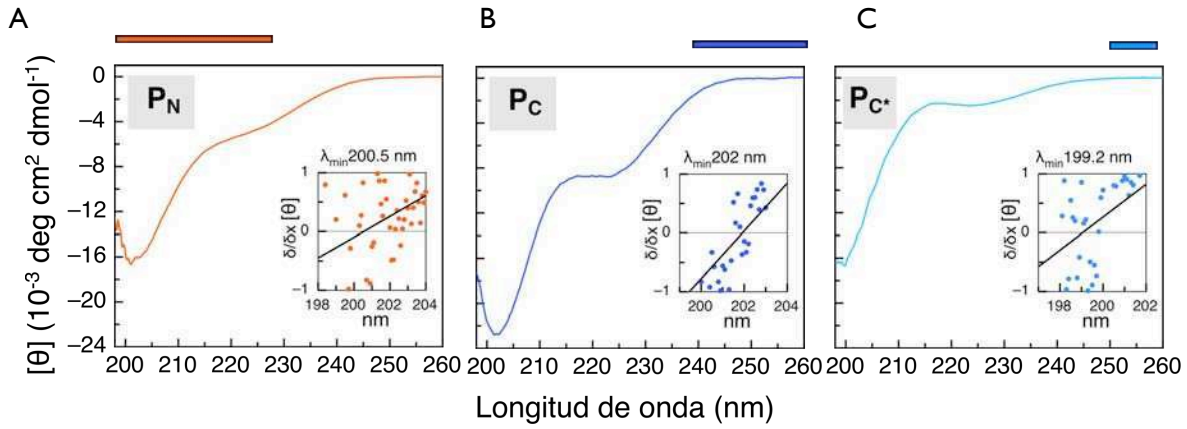


Figura II.6. Análisis del contenido de estructura secundaria por CD al UV lejano. Espectros de CD de 20 µM de los módulos monoméricos P_N (A), P_C (B) y P_{C*} (C). Recuadro: determinación del mínimo de CD. Primera derivada de los espectros de CD. La posición del mínimo corresponde al punto donde la curva es igual a cero. Todos los espectros fueron medidos en buffer 20 mM Tris.Cl pH 7,5 y 50 mM NaCl a 25 ºC.

II.2.2- Análisis del comportamiento modular de P.

Para evaluar si los distintos módulos de P son estructuralmente independientes, se trabajó con los pares complementarios de P: (P_N y P_{TETC}) o (P_{NTET} y P_C). Si existe interacción entre los distintos módulos dentro de la misma proteína, el espectro de la suma teórica debería diferir del espectro mezcla, evidenciando la interacción. Por otra parte, si los espectros de CD de la mezcla de cantidades estequiométricas de los distintos pares complementarios son idénticos al espectro de CD de la proteína P completa, implica que los módulos presentan la misma conformación en estado aislado que en el contexto de la proteína, indicando a su vez que son fragmentos estructuralmente independientes.

En primer lugar, no se encontraron diferencias entre los espectros de CD de las distintas mezclas de fragmentos complementarios (P_{NTET} y P_C) o (P_{TETC} y P_N) con la suma aritmética de los espectros individuales (Figura II.7-A y B, líneas verdes y rojas), indicando que no hay interacción entre los dominios representados por dichos fragmentos en el contexto de la proteína.

A continuación, comparamos los espectros de CD de las mezclas P_N-P_{TETC} (Figura II.7-A) o P_{NTET}-P_C (Figura II.7-B) con el espectro de la fosfoproteína P completa, a una concentración de 5 µM. Los espectros de CD de la mezcla P_N-P_{TETC} resultaron idénticos al espectro de P (Figura II.7-A, línea roja y línea negra), indicando que los módulos P_N y P_{TETC} son estructuralmente independientes. Sin embargo, al comparar el espectro de CD obtenido para la mezcla P_{NTET}-P_C con el espectro de P, los mismos resultaron significativamente diferentes (Figura II.7-B, linea roja y negra), debido al leve corrimiento del mínimo de 205 a 206 nm y por un aumento en la intensidad de la banda a 220 nm en la proteína completa.

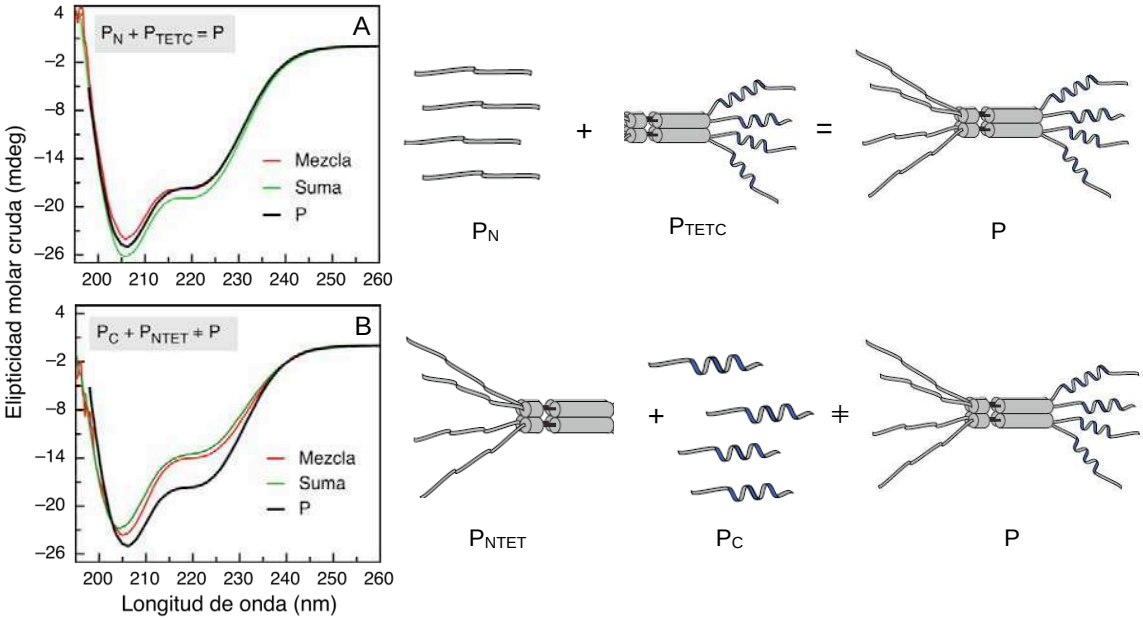


Figura II.7. Análisis del comportamiento modular de P. A-Espectro de CD al UV-lejano de P (línea negra), de la mezcla de P_N y P_{TETC} (línea roja), y de la suma aritmética de los espectros individuales de P_N y P_{TETC} (línea verde). B-Espectro de CD al UV-lejano de la P (línea negra), mezcla de P_C y P_{NTET} (línea roja), y la suma aritmética de los espectros individuales de P_C y P_{NTET} (línea verde). Todas las medidas fueron realizadas en buffer 20 mM Tris-HCl (pH7.5), 50 mM NaCl en una concentración de proteína de 5 µM, a 25 ºC.

Este resultado indica que existe un aumento en el contenido de α-hélice de P cuando el módulo C-terminal se encuentra en el contexto de la proteína completa, y demuestra que el dominio C-terminal no es completamente independiente del resto de la proteína.

II.2.3- Análisis de cambios estructurales inducidos por temperatura.

Estudios al equilibrio. Uno de los abordajes que puede ser utilizado para estudiar cuantitativamente la estabilidad de una proteína es el análisis de su estado de plegamiento con la temperatura al equilibrio. Las IDPs en particular, presentan un comportamiento térmico característico con transiciones térmicas reversibles y poco cooperativas [23].

A fin de estudiar el comportamiento térmico de P, se midieron espectros de CD a 5 °C, 25 °C, 37 °C y 85 °C, y las curvas de desnaturalización térmica entre 5 y 85 °C siguiendo el cambio de

la banda a 222 nm, para P y los distintos fragmentos. Para las curvas de desnaturalización, se probaron dos velocidades de barrido de temperatura, 1 °C/min y 5 °C/min, no observando diferencias en la forma de las curvas medidas, lo cual indica que la reacción de desplegado térmico evaluada alcanza el equilibrio en ambas velocidades ensayadas.

En la figura II.8 se muestran las curvas de desnaturalización para cada fragmento tetramérico. Decidimos presentar cada una en una escala óptima para facilitar la observación de cada transición térmica.

El contenido de estructura secundaria de P es sensible a variaciones en la temperatura siendo dicha transición completamente reversible (Figura II.8-A). Entre 5 °C y 37 °C existe pérdida de señal a 220 nm, con un ligero corrimiento del mínimo a 206 nm hacia longitudes de onda menores, sugiriendo el desplegado de un elemento de estructura marginalmente estable (Figura II.8-A). A 85 °C el espectro de CD de P aún presenta estructura secundaria significativa, muy probablemente perteneciente al dominio *coiled-coil* de tetramerización (Figura II.8-A, línea roja). La curva de desnaturalización térmica de P presenta dos transiciones estructurales: una transición entre 5 y 60 °C, y una segunda transición a partir de 65 °C que no se completa en el rango de temperatura estudiado. La primera transición corresponde a la pérdida de elementos de estructura marginalmente estable observada en los espectros de CD de P (Figura II.8-A, curva azul y celeste), mientras que la segunda transición se encuentra asociada posiblemente al desplegado y disociación del dominio de tetramerización P_Y (Figura II.8-A).

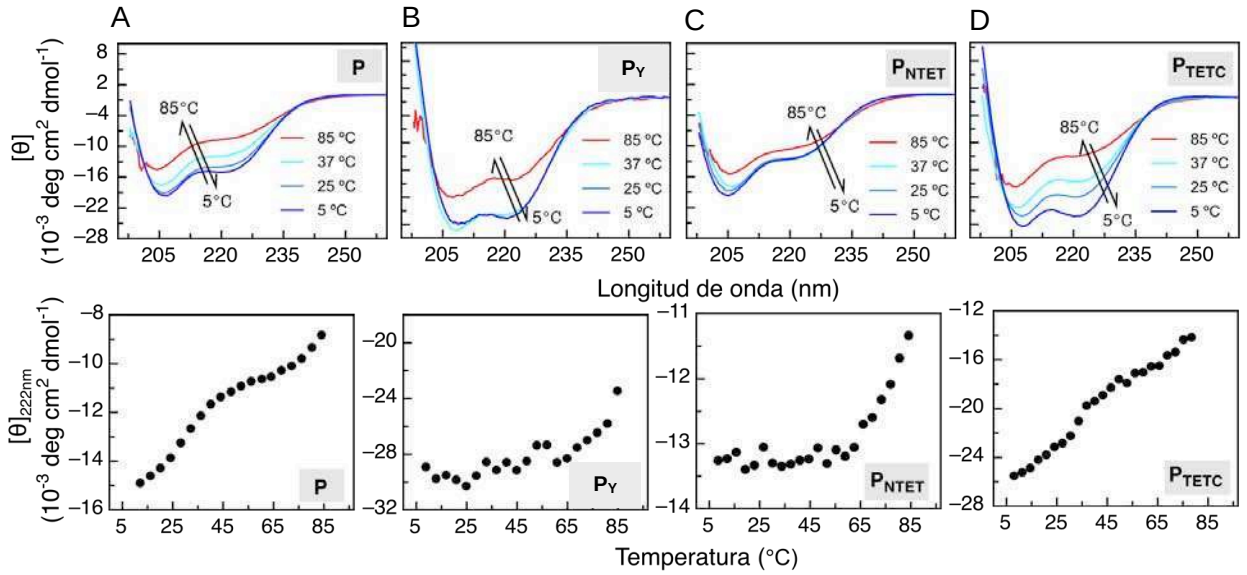


Figura II.8. Análisis de la estabilidad térmica de P. Panel superior: espectros de CD para los distintos fragmentos medidos a 5, 25, 37 y 85 ºC. En el panel inferior: curvas de desnaturalización térmica medidas entre 5 y 85 ºC siguiendo el cambio en la banda de CD a 222 nm con la temperatura, a 5 ºC/min. Todas las medidas fueron realizadas en 20 mM Tris-HCl (pH7.5), 50 mM NaCl en una concentración de proteína de 10 µM.

En contraste con lo que se observa en P, el dominio de tetramerización P_Y, no presenta cambios globales de estructura secundaria entre 5 y 60 °C. Sin embargo, se puede observar el

comienzo de la segunda transición de desplegado (Figura II.8-B), demostrando que la misma se debe a la desnaturalización y disociación del dominio de tetramerización P_Y.

El fragmento P_{NTET} posee un comportamiento térmico similar al observado para P_Y, presentando únicamente la segunda transición de desplegado térmico (Figura II.8-C). En contraste, el fragmento P_{TETC}, posee la misma transición estructural a baja temperatura observada para P (Figura II.8-D), con una pérdida del contenido de estructura α-hélice entre 5 y 37 °C (Figura II.8-D, curva azul y celeste). Esto sugiere que el elemento de estructura marginalmente estable identificado previamente en P (Figura II.8-A) se encuentra localizado en el dominio C terminal.

Por otra parte, se estudiaron los fragmentos monoméricos correspondientes a los dominios N y C terminal de P. El fragmento P_N, presenta un espectro de CD típico de una proteína o región intrínsecamente desordenada, con una transición térmica no cooperativa y reversible (Figura II. 9-A). Se puede observar que el aumento de la temperatura induce una disminución de la bandas a 220 nm y un aumento en el mínimo a 200 nm con un ligero corrimiento hacia longitudes de onda mayores (Figura II.9-A, panel superior). Al analizar el espectro diferencia entre 5 y 85 °C, observamos que el mismo presenta un componente positivo a 218 nm y con un mínimo pronunciado a 200 nm (Figura II.9-A, panel medio), lo que podría sugerir estabilización de estructura PII a bajas temperaturas [30].

En contraste, al analizar el comportamiento térmico del módulo P_C, observamos una mayor estabilización de estructura secundaria a bajas temperaturas (Figura II.9-B, panel superior). Al analizar el espectro diferencia entre 5 y 85 °C, se observa que la estructura que se pierde durante la transición térmica presenta características de α-hélice, con mínimos a 208 y 220 nm (Figura II.9-B, panel medio). Por otro lado, se puede observar que la transición térmica del fragmento evaluada a 222 nm presenta cierta cooperatividad (Figura II.9-B, panel inferior). Este resultado indica, el elemento estructural marginalmente estable identificado en el dominio C-terminal (Figura II.8-D) también es estabilizado en el fragmento P_C correspondiente al módulo C-terminal aislado.

Es interesante destacar que dicha transición térmica no se puede apreciar en P_{C*} (Figura II.9-C). El sub-fragmento P_{C*} se comporta como un fragmento desordenado en solución (Figura II.9-C), con un comportamiento térmico similar al observado para P_N. Al analizar el espectro diferencia de este fragmento podemos observar una banda positiva a 218 nm y un mínimo a 200 nm, indicando que los últimos 44 residuos de P_C pueden explorar conformaciones tipo PII.

En conclusión, estos resultados indican que el fragmento P_C contiene al elemento marginalmente estable presente en el dominio C-terminal de P (P_{TETC}) y sugieren que el mismo se localiza dentro de los primeros residuos del módulo C terminal, o bien que se requiere de toda la información de secuencia del módulo completo para estabilizar su estructura.

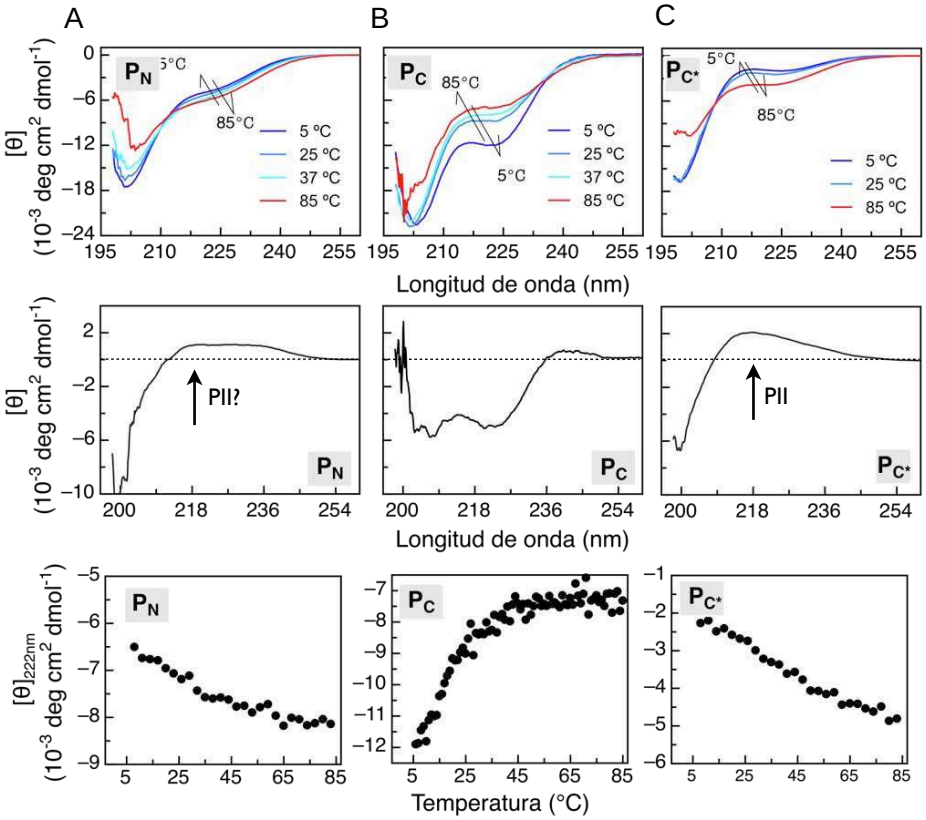


Figura II.9. Análisis de la estabilidad térmica de los fragmentos monoméricos de P. Determinaciones realizadas para P_N (A), P_C (B) y P_{C*} (C). Panel superior: espectros de CD medidos a 5, 25, 37 y 85 ºC. Panel medio: espectro diferencia de CD entre 5 y 85 ºC. En el panel inferior: curvas de desnaturalización térmica medidas entre 5 y 85 ºC siguiendo el cambio en la banda de CD a 222 nm con la temperatura. Todas las medidas fueron realizadas en 20 mM Tris-HCl (pH7.5), 50 mM NaCl a una concentración de 15 µM.

Con la finalidad de evaluar cómo el contexto influye la estabilidad del elemento de estructura marginalmente estable, se ajustaron los resultados de la primera transición térmica observada entre 5 y 60 °C a la ecuación de Gibbs-Helmholtz para las distintas especies conteniendo dicho elemento (ver materiales y métodos) (Figura II.10).

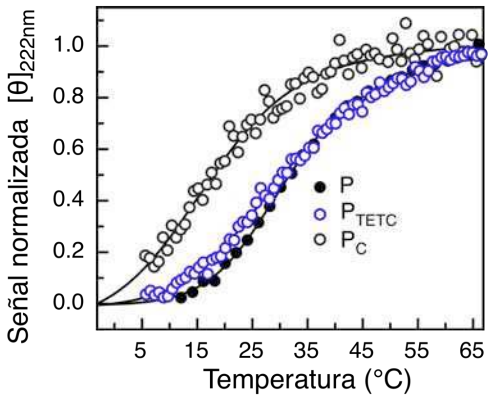


Figura II.10. Análisis de la primera transición de desnaturalización térmica de P. Curvas de desnaturalización térmica normalizadas entre 0 y 1, monitoreando la banda a 222 nm para P (círculos negros), P_{TETC} (círculos azules abiertos) y P_C (círculos negros abiertos). Los puntos medios de las transiciones térmicas (*Tm*) para P, P_{TETC} y P_C fueron de 26 ± 2 ºC, 25 ± 1 ºC, 15 ± 2 ºC, respectivamente. El valor de la *Tm* representa el promedio de tres medidas experimentales individuales.

A partir de dicho ajuste determinamos los valores de los puntos medios de la transición térmica (T_m) para P, P_{TETC} y P_C. En la figura II.10, se muestran los datos de las primeras transiciones normalizados entre 0 y 1, a fin de comparar entre fragmentos. Las transiciones de temperatura entre P y P_{TETC} son idénticas, presentando valores de T_m estimada de 26 ± 2 °C y 25 ± 1 °C, respectivamente; mientras que el fragmento P_C presenta una T_m de 15 ± 2 °C. Esta diferencia de 10 °C entre las T_m, demuestra que el elemento marginalmente estable en el contexto del dominio de tetramerización posee mayor estabilidad en comparación con el módulo C-terminal aislado.

Estudios cinéticos de la primera transición de desplegado. Se estudió la cinética del cambio estructural correspondiente a la primera transición térmica para P y el módulo P_C. Para tal fin se equilibró la proteína en un baño termostatizado a 10 o 40 °C, para luego ser transferida a una cubeta con buffer equilibrado a 40 o 10 °C, según el caso. Las medidas fueron realizadas en cubetas de distinto paso óptico (0.2, 0.5 y 1 cm) para descartar posibles artefactos debido a la mezcla de dos volúmenes de muestra (buffer y proteína) equilibrados a temperaturas distintas. Se evaluó el cambio en la señal de la banda de CD a 222 nm en el tiempo, donde se esperaba obtener un cambio gradual de la señal hasta llegar al equilibrio.

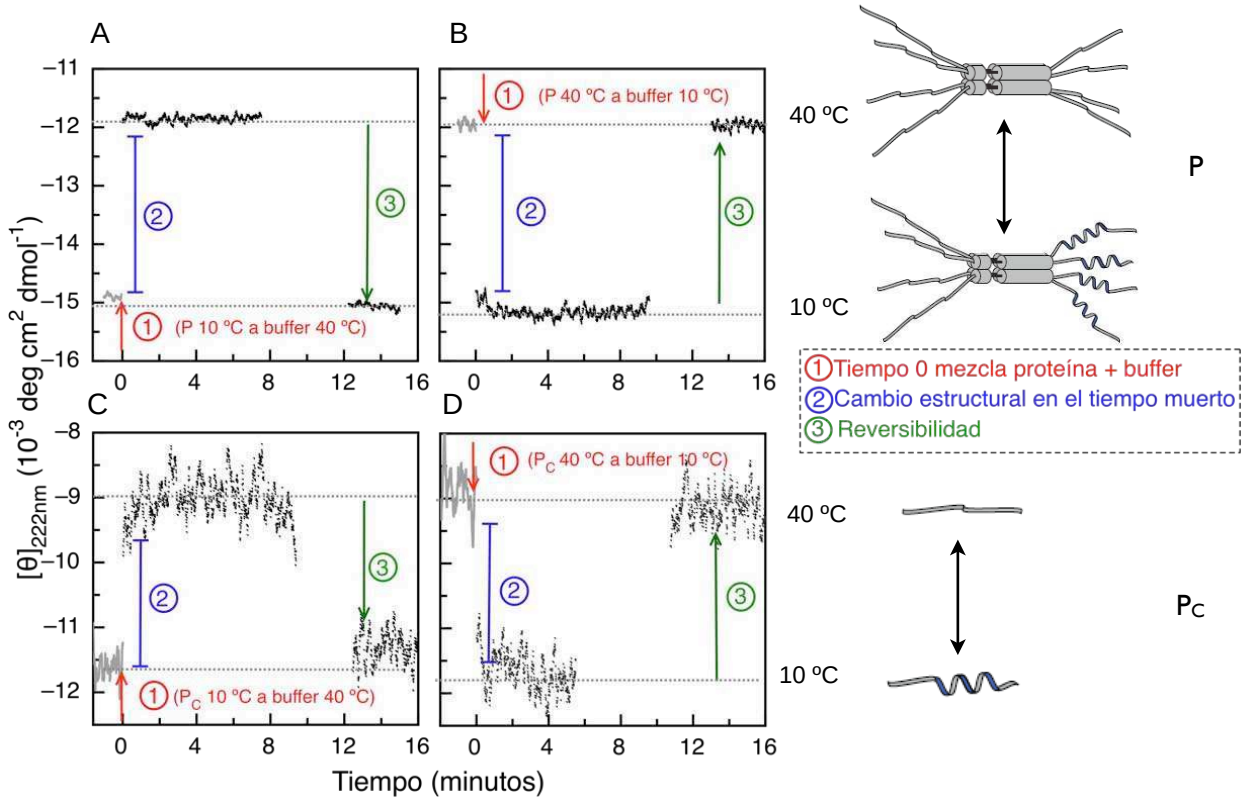


Figura II.11. Análisis cinéticos de la primera transición de desplegado por temperatura. Un stock concentrado de P (A) o P_C (C) fue incubado a 10 ºC por 15 minutos y fue transferido a buffer equilibrado a 40 ºC (flecha roja). El stock de P (B) y PC (D) fue incubado a 40 ºC por 15 minutos y fue transferido a buffer equilibrado a 10 ºC (flecha roja). Todas las medidas fueron realizadas en buffer 20 mM Tris-HCl (pH7.5), 50 mM NaCl, a una concentración final en cubeta de 6 µM.

Notablemente, los cambios estructurales inducidos por temperatura para ambas especies P y P_C ocurrieron dentro del tiempo de mezclado (menor a 10 segundos) (Figura II.11-A y B, línea azul), lo cual demuestra que la plegado-desplegado del elemento marginalmente estable del domino C-terminal ocurre con una cinética rápida. Para todos los casos evaluados el cambio estructural fue reversible, ya que se recuperaron los valores de elipticidades iniciales luego de realizar el cambio de temperatura (Figura II.11-A y B, flecha verde).

II.2.4- Estudios de la estabilidad química de P.

Estudios al equilibrio. En un trabajo previo de nuestro laboratorio, se demostró que la fosfoproteína P presenta una transición de desplegado químico de tres estados [132]. La primera transición reversible ocurre entre 0 y 3.5 M Gdm.Cl donde el tetrámero permanece estable, y la segunda transición ocurre entre 4 y 6 M Gdm.Cl que corresponde al desplegado y disociación del tetrámero, a partir del cual se estimó una constante de disociación y desplegado de 10^{-28} M^3 (Figura II.12) [132]. En esta tesis nos centramos en el estudio de la primer transición de desplegado, que se completa a concentraciones bajas de Gdm.Cl, pues proponemos que corresponde al desplegado de elementos de estructura marginalmente estable presente en el dominio C terminal de P.

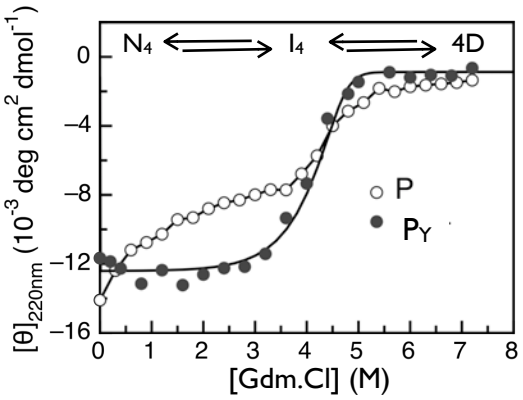


Figura II.12. Desnaturalización química de P con Gdm.Cl. Figura adaptada de [132]. Se muestran las transiciones de desplegado con Gdm.Cl de P (círculos abiertos) y de P_Y (círculos grises).

Para determinar la identidad de la región de P donde se localiza el/los componentes de estructura responsable de la primera transición de desplegado, se realizaron titulaciones con concentraciones crecientes de Gdm.Cl en un rango entre 0 y 3.5 M Gdm.Cl, a 20 °C y a 10 °C, incubando las muestras 5 minutos entre mediciones, tiempo en el que el sistema llega al equilibrio. La pérdida de estructura secundaria fue seguida por CD al UV lejano, evaluando la disminución gradual de la señal de CD con el aumento de la concentración de Gdm.Cl.

En primer lugar, se estudió la primera transición de desplegado para los fragmentos tetraméricos P_{TETC} y P_{NTET} en comparación con P. Se puede apreciar que para P_{TETC}, a medida

que aumenta la concentración del agente desnaturalizante, las señales espectrales reportan el estado desnaturalizado de la primera transición de desplegado de la proteína, donde la elipticidad molar a 220 nm varía desde -20 mdeg hasta -10 deg entre 0 y 3.5 M Gdm.Cl (Figura II.13-A). Por otro lado, las señales espectrales correspondientes a la transición de desplegado del fragmento P_{NTET} presentan una menor diferencia entre 0 y 3.5 M Gdm.Cl, variando entre -12 mdeg y -10 mdeg entre 0 y 3.5 M Gdm.Cl, respectivamente (Figura II.13-B). Esto indica, de forma consistente con los resultados anteriores, que P_{TETC} contiene una mayor proporción de estructura secundaria/terciaria global que el fragmento P_{NTET}.

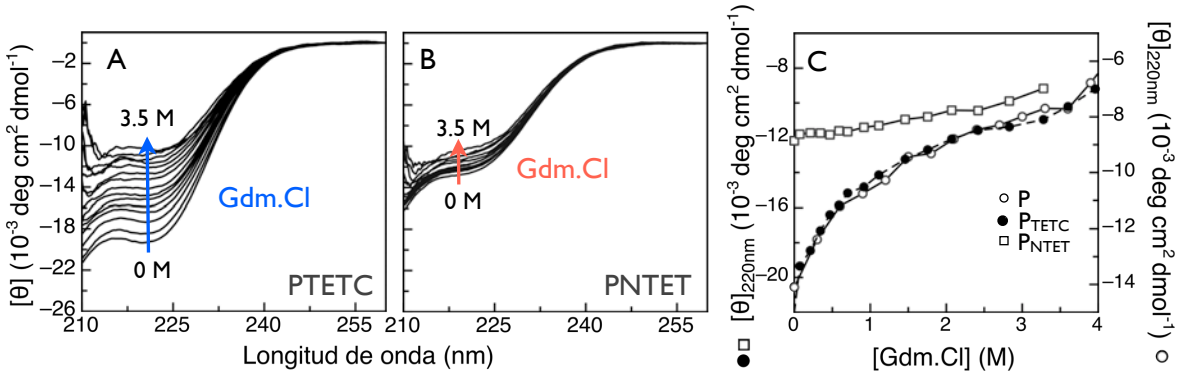


Figura II.13. Análisis de la primera transición de desplegado químico de P. A y B-Experimentos de titulación con Gdm.Cl (0 a 3.5 M) seguidos por CD al UV lejano para P_{TETC} (A) y para P_{NTET} (B). La flecha indica el sentido de la reacción. C-Análisis del desplegado con Gdm.Cl evaluando la posición de la banda a 220 nm a distintas concentraciones de Gdm.Cl para P_{TETC} (círculos negros) y P_{NTET} (cuadrados blancos), en comparación con P (círculos blancos). Todos los experimentos fueron realizados por titulación, en buffer 20 mM Tris.Cl pH 7.5, 50 mM NaCl y a una concentración de proteína de 10 µM, a 20 ºC.

El análisis del cambio de los valores de elipticidad de la banda de CD a 220 nm a distintas concentraciones de Gdm.Cl, muestra que las especies P y P_{TETC} presentan curvas de desplegado idénticas, las cuales presentan cierto grado de cooperatividad (Figura II.13 C). Por otra parte, P_{NTET} presenta una transición no cooperativa y de menor magnitud comparada con el cambio global observado en P_{TETC} (Figura II.13 C). Estos resultados indican que P_{TETC} contiene a todos los elementos de estructura secundaria/terciaria que se pierden durante la primera transición de desplegado de P, incluyendo aquellos que se despliegan en P_{NTET}.

Para estimar el aporte global de cada módulo a la primera transición de desplegado de P, se evaluaron las magnitudes del cambio espectral luego del desplegado químico en 3.5 M Gdm.Cl, analizando los datos crudos de elipticidad de las muestras medidas a 10 µM para poder comparar entre fragmentos con elipticidades molares distintas ([θ]). Por otra parte, para estabilizar el componente de estructura marginalmente estable presente en el domino C terminal de P, estas medidas fueron realizadas a 10 °C.

En la figura II.14 podemos observar que para las especies P y P_{TETC}, que las magnitudes globales de cambio en estructura secundaria son similares (20 mdeg de cambio). Por otro lado, P_{NTET} presenta un cambio de señal de 7 mdeg, el cual corresponde al 35% del cambio total

observado para P. Esta observación sugiere la existencia de al menos dos elementos estructurales que se despliegan en P durante la primera transición con Gdm.Cl.

Por lo tanto, dado que P y P_{TETC} presentan el mismo cambio global durante transición con Gdm.Cl, y que el dominio de tetramerización P_Y es estable a 3.5 M Gdm.Cl [132], proponemos que las transiciones observadas corresponden al desplegado del dominio C terminal y al módulo P_Z de la región central de P (P_{TET}) (Figura II.14, esquema).

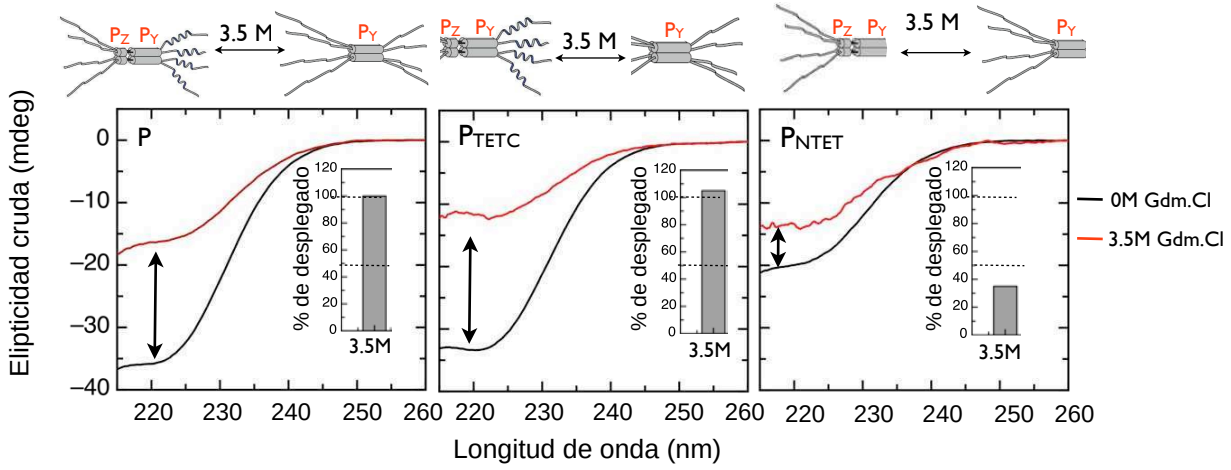


Figura II.14. Cambio global de la señal de CD a 3.5 M Gdm.Cl. Cambio en la señal de CD a 220 nm con el agregado de 3.5 M Gd.Cl para P, P_{TETC} y P_{NTET}. En el interior de cada gráfico se indica el % del cambio relativizado a la primer transición de desplegado de P. Todos los experimentos fueron realizados en buffer 20 mM Tris.Cl pH 7.5, 50 mM NaCl y a una concentración de proteína de 10 µM, a 10 ºC.

Para estudiar la estabilidad de dichos elementos estructurales, analizamos el cambio en la señal de CD a concentraciones bajas de Gdm.Cl (0.5 M y 1M) (Figura II.15).

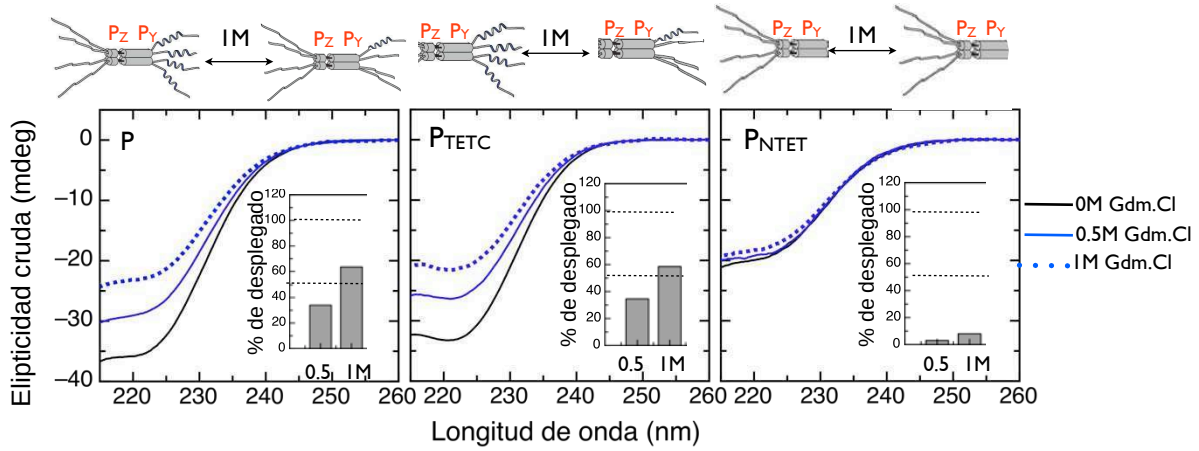


Figura II.15. Cambio global de la señal de CD a 0.5 y 1M Gdm.Cl. Cambio de la señal de CD con el agregado de 0.5M Gdm.Cl (línea azul) y 1 M Gdm.Cl (línea punteada). En el interior de cada gráfico se indica el % del relativizado a la primer transición de desplegado de P. Todos los experimentos fueron realizados en buffer 20 mM Tris.Cl pH 7.5, 50 mM NaCl y a una concentración de proteína de 10 µM, a 10 ºC.

Observamos que a ambas concentraciones, las únicas especies que presentan pérdidas significativas de la señal a 220 nm son P y P_{TETC}, siendo dicho cambio idéntico en magnitud

para ambas especies (Figura II.15). A 0.5 M Gdm.Cl, P y P_{TETC} presentan una pérdida del 30% de la señal correspondiente a la primera transición de desplegado, mientras que a 1M Gdm.Cl observamos una pérdida del 60% de la misma (Figura II.15). En cambio, para P_{NTET} no se aprecian cambios en la estructura secundaria en estas condiciones (Figura II.15). Por lo tanto, el elemento marginalmente estable del dominio C-terminal posee una estabilidad menor en comparación con el elemento estructural localizado en el módulo P_Z, el cual requiere concentraciones superiores a 1 M Gdm.Cl para desplegarse.

Cinética del desplegado por Gdm.Cl. Para analizar en forma comparativa la velocidad con la que los componentes de estructura secundaria/terciaria de P se forman o se rompen, se evaluó la cinética del cambio estructural siguiendo el cambio de la señal de CD a 220 nm para P y P_{NTET} mediante agregado de distintas concentraciones de Gdm.Cl, en idénticas condiciones de buffer que las medidas al equilibrio. Se evaluó la primera transición de desplegado de P, mediante el agregado de 0.5 M y 3.5 M Gdm.Cl (Figura II.16-A).

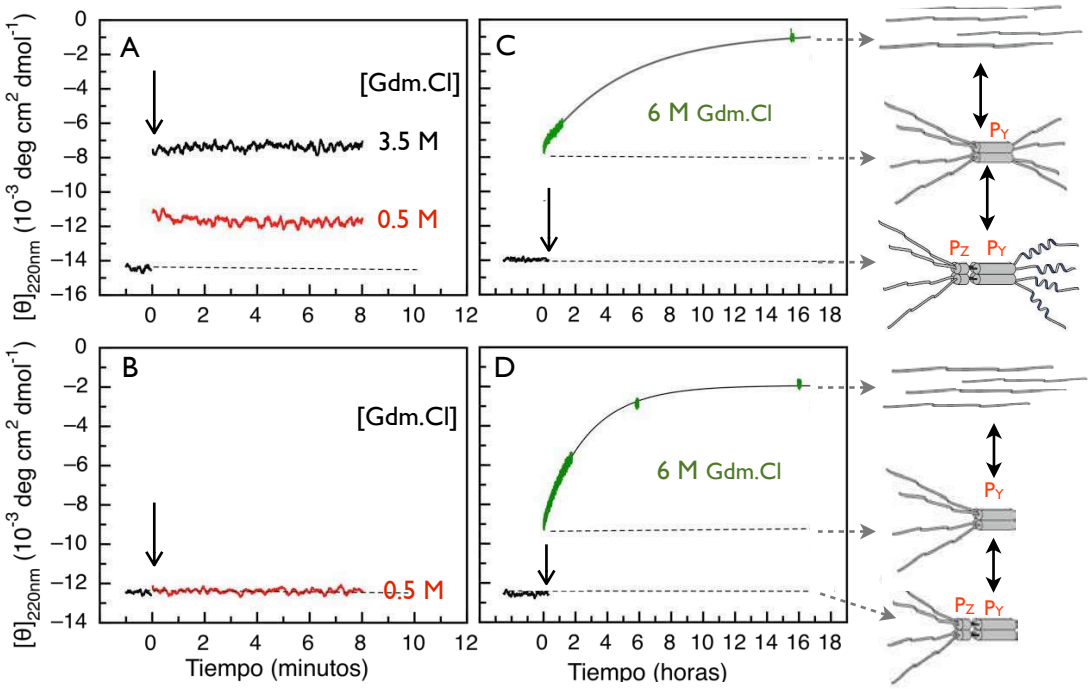


Figura II.16. Análisis cinético de las transiciones de desplegado químico. Se evaluó el cambio en la estructura secundaria al agregar 0.5 M (línea roja), 3.5 M (línea negra) y 6M Gdm.Cl para P (A y C) y para P_{NTET} (B y D). La segunda transición de desplegado fue ajustado a un modelo de función exponencial simple obteniendo una k=0.0032 min^{-1} para P y k=0.007 min^{-1} para P_{NTET}, las cuales corresponden a un $t_{1/2}$ de ~ 4 y ~ 2 horas, respectivamente. La flecha negra indica el momento en donde se adiciona Gdm.Cl. Todas las medidas fueron realizadas en 20 mM Tris-HCl (pH7.5), 50 mM NaCl a 10 ºC, a una concentración de 10 µM.

Al igual que lo observado en los experimentos con temperatura, el desplegado de P a ambas concentraciones ocurre durante el tiempo de mezcla (< 10 segundos) (Figura II.16-A). Este resultado demuestra que el desplegado del elemento de estructura del dominio C terminal y del módulo P_Z, presentan una rápida inter-conversión entre la especie estructurada y la especie

desplegada. Por otra parte, al igual que en los experimentos al equilibrio, no se observan cambios en la estructura secundaria de P_{NTET} al agregar 0.5 M Gdm.Cl (Figura II.16-B).

En la segunda transición de desplegado, estudiada en P y $P_{NTET,}$ nuevamente se observa el cambio conformacional rápido correspondiente a la primera transición de desplegado (Figura II. 16-C y D, primer salto). Además, se puede apreciar una segunda transición de desplegado (Figura II.16-C y D, línea verde), la cual corresponde al desplegado y disociación del dominio de tetramerización y requiere varias horas para su conclusión ($t_{1/2} \geq 2$ horas).

II.2.5- Unión a la sonda fluorescente 8-anilino-1-naftalenosulfonato (ANS).

La sonda fluorescente ANS es un reportero sensible utilizado para identificar zonas hidrofóbicas accesibles al solvente en proteínas, aunque con cierto grado de estructura terciaria. Por esta propiedad, el ANS se puede utilizar para caracterizar estados parcialmente desplegados tipo *molten globule* [133]. Por este motivo, ha sido utilizada como herramienta para estudiar o identificar distintos grados de desorden en IDPs [134, 135]. En particular, en un estudio realizado sobre la proteína intrínsecamente desordenada clusterina, se determinó la presencia de una región tipo *molten globule* analizando su capacidad de unir ANS en distintas concentraciones de urea [135].

El ANS en solución acuosa presenta un espectro de emisión de fluorescencia con un máximo alrededor de 520 nm cuando es excitado a 370 nm (Figura II.17-A). Sin embargo, al interactuar con una proteína en estado tipo *molten globule* o por interacción con sitios o superficies hidrofóbicas de proteínas globulares, se produce un aumento significativo en el espectro de emisión de fluorescencia con un máximo alrededor 475 nm (Figura II.17-A).

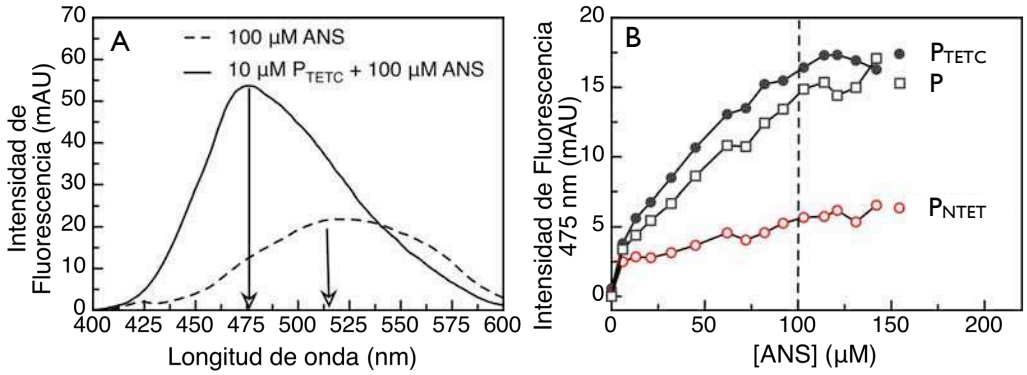


Figura II.17. Análisis de la unión a ANS. A-Espectro de emisión de fluorescencia de 100 µM ANS (línea punteadas) y 10 µM P_{TETC} incubado con 100 µM ANS por 30 minutos en oscuridad (línea llena). Las muestras fueron excitadas a 370 nm y los espectros de emisión registrados entre 400 y 600 nm. Las flechas indican la posición de los máximos de intensidad de fluorescencia. B-Titulaciones con ANS para 5 µM de P, P_{TETC} y P_{NTET}, siguiendo el cambio de la intensidad de fluorescencia a 475 nm. Dicha longitud de onda representa la longitud de onda máxima para todos los complejos ANS:P. Todas las medidas fueron realizadas en buffer 20 mM Tris-HCl (pH7.5), 50 mM NaCl y a 20 ℃.

Nuestro grupo demostró que P es capaz de unir ANS y que dicha interacción se pierde en presencia de 2 M Gdm.Cl [132], lo cual sugiere que dicha unión reportaría la presencia del elemento estructural marginalmente estable del dominio C-terminal de P.

Para seleccionar una condición de trabajo que nos permita comparar los distintos fragmentos, se titularon P, P_{TETC} y P_{NTET} con concentraciones crecientes de ANS, y se evaluó el aumento de la señal de intensidad de fluorescencia a 475 nm (Figura II.17-B). La curva de unión para P y P_{TETC} presenta una forma similar, donde se observa que los sitios de unión a ANS se encuentran saturados a 100 µM (Figura II.17-B, línea punteada), concentración que se eligió posteriormente para el análisis de los distintos fragmentos. Por otra parte, P_{NTET} une ANS en menor proporción que la observada para la proteína completa y no llega a saturación en las concentraciones evaluadas (Figura II.17-B).

Se evaluó en forma comparativa la capacidad de unir ANS para los distintos fragmentos monoméricos y tetraméricos de P (Figura II.18-A). En el gráfico de barras se aprecia que P y P_{TETC} unen ANS de manera similar, mientras que P_{NTET} lo une en una menor proporción (Figura II.18-A). Sin embargo, ni el dominio de tetramerización P_Y, ni los fragmentos monoméricos P_N y P_C presentan afinidad por la sonda (Figura II.18-A). El hecho de que en los estados aislados P_Y y el módulo C-terminal P_C no puedan unir ANS indican que el contexto tetramérico del domino C-terminal es necesario para que se estabilice un plegamiento terciario accesible al solvente que de lugar a la unión de ANS.

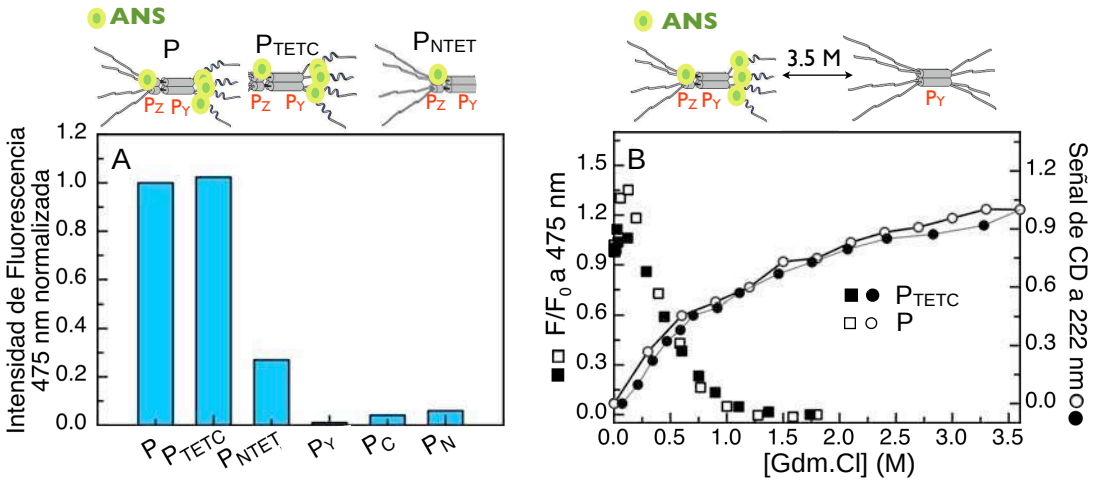


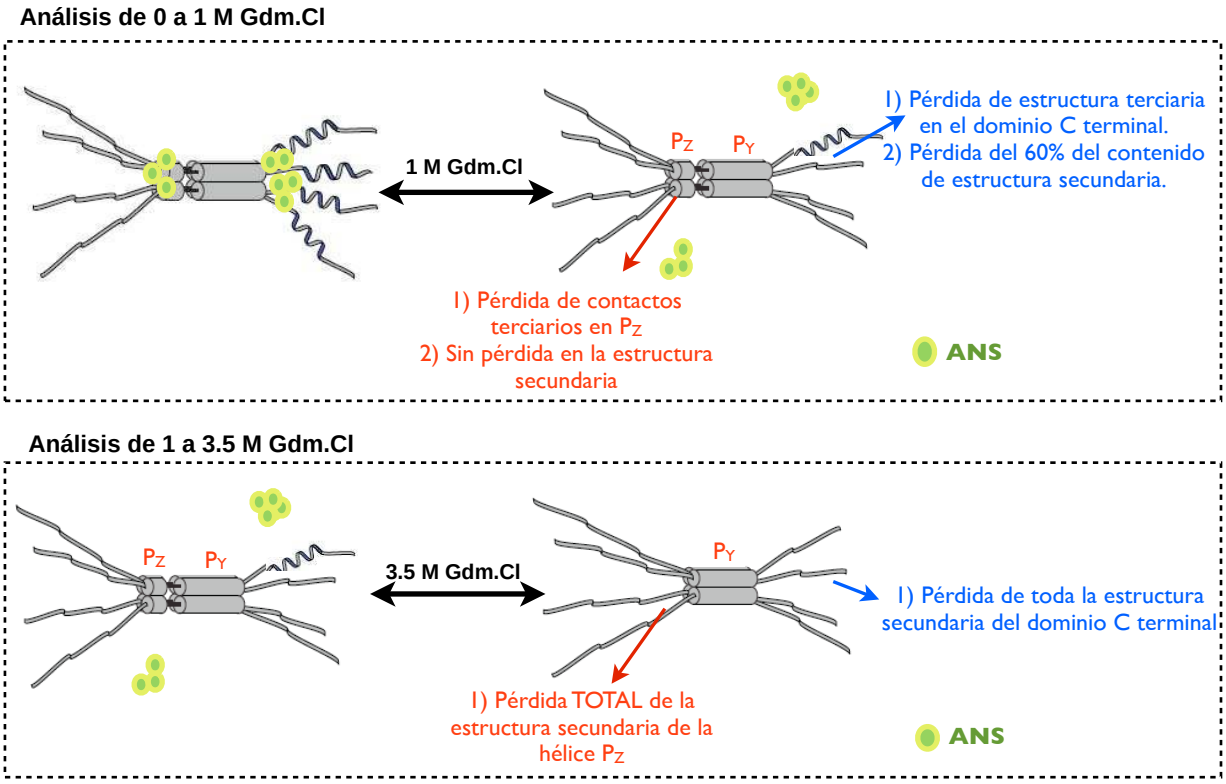
Figura II.18. Análisis de la unión a ANS por parte de P y los distintos fragmentos. A-Disección de la capacidad de unir ANS de los distintos fragmentos siguiendo el cambio en la intensidad de fluorescencia a 475 nm. Los resultados están expresados como fracción de la capacidad de unir ANS normalizada a la unión de la proteína P entera. B-Desplazamiento de ANS con concentraciones crecientes de Gdm.Cl, seguidos por el cambio en la intensidad de fluorescencia a 475 nm (eje izquierdo) y por el cambio en la señal de CD a 220 nm (eje derecho). Las titulaciones fueron realizadas incubando 5 minutos entre punto y punto, en buffer 20 mM Tris-HCl (pH7.5), 50 mM NaCl más 100 µM ANS, a una concentración final en cubeta de 10 µM, a 20 ºC.

Por otro lado, con el fin de analizar la estabilidad del o de los sitios de unión a ANS presentes en P, se evaluó la pérdida de la señal a 475 nm al titular con concentraciones crecientes de Gdm.Cl las especies P y P_{TETC} en complejo con 100 µM ANS. Dichos datos fueron analizados

en conjunto con el cambio en estructura secundaria seguido por CD en las mismas condiciones.

Las curvas de desplazamiento de ANS obtenidas para P y P_{TETC} resultaron idénticas (Figura II. 18-B). En ambas especies el ANS unido fue completamente desplazado a 1M Gdm.Cl (Figura II.18-B), mientras que a esta concentración el cambio en estructura secundaria corresponde al 60% del total de la transición de desplegado (Figura II.18-B).

Si integramos los resultados de los experimentos de ANS con el análisis de desplegado con Gdm.Cl del apartado anterior, podemos identificar al menos dos módulos que se ven afectados dentro de la primera transición de desplegado de P con Gdm.Cl: la hélice P_Z de la región central de P, y el dominio C terminal de P. A continuación se muestra un esquema que resume los cambios experimentados por P durante la primer transición de desplegado con Gdm.Cl (Esquema I).



Esquema I. Resumen de los cambios estructurales experimentados por P durante la primera transición de desplegado con Gdm.Cl.

En primer lugar, a partir de los experimentos de desplegado químico se observa que entre 0 y 1 M Gdm.Cl ocurre la pérdida del 60% de estructura secundaria del domino C terminal de P, mientras que la estructura secundaria de P_{NTET} no es afectada (Figura II.15). Sin embargo, demostramos que a 1 M Gdm.Cl se pierde toda la estructura terciaria que une ANS en P (Figura II.18-B). Por otro lado, a concentraciones mayores a 1 M Gdm.Cl se terminan de

desplegar el elemento marginalmente estable del dominio C terminal de P, así como también se pierde el 100% de la estructura secundaria de la primera hélice de P_{TET} (Figura II.14).

Estos resultados nos permiten concluir que el elemento marginalmente estable presente dentro del dominio C-terminal contiene estructura secundaria y contactos involucrados en la formación de estructura terciaria. Por lo tanto, el dominio C terminal de P presenta estructura marginalmente estable, flexible y dinámica con características de tipo *pre-molten-globule*.

II.3. TRANSICIONES ESTRUCTURALES Y ESTUDIOS CONFORMACIONALES DE LOS FRAGMENTOS P_N Y P_C.

II.3.1- Estabilización estructural inducida por pH.

Se ha propuesto que la repulsión electrostática entre residuos con igual carga podría explicar al menos en parte la ausencia de estructura definida en IDPs a pH neutro [118]. Un abordaje utilizado para el estudio de estructuras transitorias o débiles en equilibrio en las IDPs es analizar el efecto de la neutralización de los carboxilatos de los aminoácidos ácidos por reducción en el pH [23]. Como ya hemos mencionado, P es una proteína ácida con un pI teórico de 4.52. Hemos observado que P y los fragmentos tetraméricos P_{NTET} y P_{TETC} agregan irreversiblemente en condiciones de pH menor a 5.0. Sin embargo, los módulos monoméricos P_N (pI= 6.58) y P_C (pI= 4.34) resultaron solubles en dichas condiciones de pH. Se estudió el efecto del pH sobre los componentes de estructura secundaria y cuaternaria de los módulos P_N y P_C, por CD y por cromatografía de exclusión molecular (Figura II.19).

Podemos apreciar que para el fragmento desordenado P_N no se observan cambios en el contenido de estructura secundaria entre pH 7.5 y pH 3.0 (Figura II.19-A). Existe una ligera compactación, si bien significativa, de su radio hidrodinámico (Figura II.19-B), pasando de un R_S= 2.5 ± 0.2 nm a pH 7.5 a un R_S= 2.3 ± 0.1 nm (Tabla II.2).

Por otra parte, para el fragmento P_C se observa que la neutralización de los carboxilatos a pH 3.0 produce un corrimiento del mínimo de 202 a 206 nm en el espectro de CD, y un aumento significativo en la intensidad de la banda a 220 nm, lo cual implica un aumento en el contenido de estructura α-hélice (Figura II.19-C). Este cambio estructural está acompañado por un retraso en el volumen de elución de la muestra en una columna de filtración molecular Superdex200 (Figura II.19-D). Dicha compactación representa un cambio significativo en el cociente MW_{SEC}/ MW_{SLS} de 3.1 ± 0.2 a 1.24 ± 0.02, similar al esperado para una proteína globular del mismo peso molecular (MW_{SEC}/MW_{SLS} ~ 1) (Figura II.19-D, panel derecho).

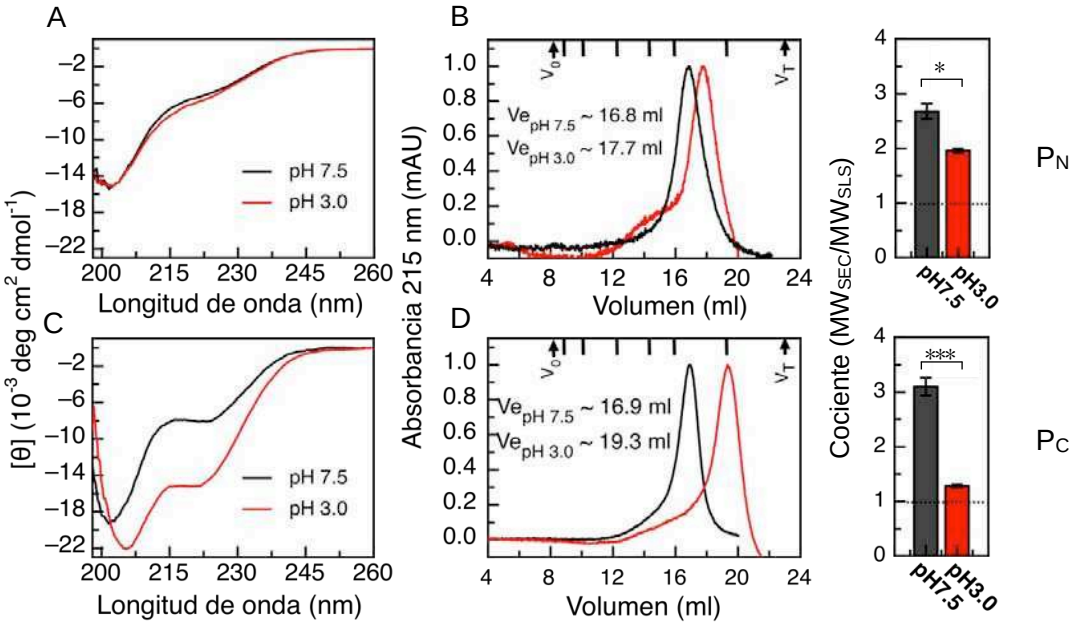


Figura II.19. Efecto del pH sobre la conformación de los módulos P_N y P_C. A y C- Espectros de CD medidos a pH 7.5 (línea negra) y a pH 3.0 (línea roja) para los fragmentos P_N (A) y P_C (C). B y D- Experimento de filtración molecular realizados a pH 7.5 (línea negra) y a pH 3.0 (línea roja) para P_N (B) y P_C (D). En la parte superior del gráfico se indican los volumenes de elución de los estándares de calibración. Las flechas indican las posiciones del V_0 y el volumen total (V_T) de la columna. Los cocientes entre el peso molecular calculado por SEC y el peso molecular calculado por SLS a pH 7.5 (barras grises) y a pH 3.0 (barras rojas), se muestran a la derecha de cada panel. La línea punteada indica el cociente esperado para una proteína globular. Las medidas fueron realizadas en buffer 20 mM Tris-HCl (pH7.5), 200 mM NaCl o 20 mM Formiato de sodio (pH 3.0), 200 mM NaCl, a temperatura ambiente. N=3. (* p< 0.05, ***p<0.001, test estadístico T de STUDENT).

Al calcular los R_S para P_C en ambas condiciones de pH, observamos que el mismo pasa de tener un R_S de 2.4 ± 0.2 nm, intermedio entre el esperado para una IDP con conformación tipo *pre molten globule* y un IDP tipo r*andom coil*, a tener un R_S de 1.8 ± 0.1 nm, intermedio entre una proteína globular y un estado tipo *molten globule* (Tabla II.2). Lo cual demuestra la marcada compactación de este módulo a bajo pH.

Tabla II.2. *Comportamiento hidrodinámico con el pH.*

	R_S SEC (nm)		**R_S Globular***	**R_S *molten globule****	**R_S IDP *pre molten globule****	**R_S IDP *random coil****	**R_S desplegada con Gdm.Cl***
	pH 7.5	*pH 3.0*					
P_N	2.5 ± 0.2	2.3 ± 0.1	1.8	2.0	2.5	2.9	3.1
P_C	2.4 ± 0.2	1.8 ± 0.1	1.6	1.9	2.3	2.5	2.7

*valores teóricos determinados utilizando las ecuaciones descriptas en [31]

Para evaluar si esta compactación se encuentra asociada a la ganancia de un plegamiento terciario capaz de interactuar con ANS, se evaluó si esta estructura compacta estabilizada a pH 3.0 presenta la capacidad de unir a la sonda (Figura II.20-A).

Como se demostró en el apartado anterior, el módulo C terminal P_C a pH 7.5 no une a la sonda fluorescente ANS (Figura II.20-A). Sin embargo, a pH 3.0 se observa unión a la sonda,

demostrado por el aumento en la intensidad de fluorescencia (Figura II.20-A). Este resultado confirma que el pH induce una transformación estructural en P_C de un estado extendido y con estructura secundaria residual, a una estructura compacta y parcialmente plegada. Además, se demostró la reversibilidad de dicha transición tras dializar la muestra a pH 3.0 contra buffer a pH 7.5 (Figura II.20-B)

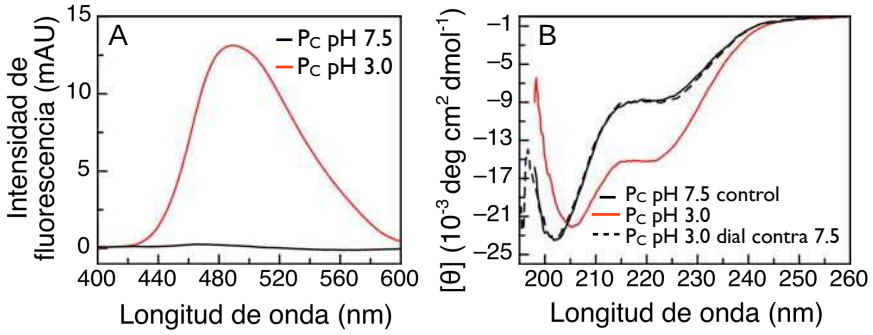


Figura II.20. Efecto del pH sobre la estructura terciaria de P_C. A-Espectro de emisión de fluorescencia de ANS para P_C medidos a pH 7.5 (linea negra) y a pH 3.0 (línea roja) luego de restarle la contribución del buffer. Las medidas fueron realizadas en buffer 20 mM Tris-HCl (pH7.5), 200 mM NaCl o 20 mM Formiato de sodio (pH 3.0), 200 mM NaCl, con 100 µM ANS. B-Espectro de CD de P_C donde se muestra la reversibilidad del cambio estructural inducido por pH.

Finalmente, se estudió si el cambio conformacional descripto para P_C se ve reflejado en el fragmento P_{C*}, correspondiente a sus últimos 44 residuos. En la figura II.21 se observa que el pH no induce estructura α-hélice en dicho fragmento, al igual que lo observado previamente para la desnaturalización térmica (Figura II.9-C). Una posible explicación es que los elementos con mayor tendencia a estabilizar hélice se encuentran dentro de los primeros residuos del dominio C terminal en la región adyacente al domino de tetramerización. No obstante, la causa más plausible es que se requiera la presencia del módulo completo para estabilizar dicha estructura.

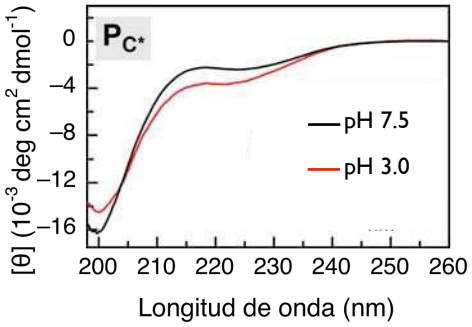


Figura II.21. Efecto del pH sobre la estructura secundaria de P_{C*}. Espectro de CD del subfragmento P_{C*} medidos a pH 7.5 (línea negra) y a pH 3.0 (línea roja). Las medidas fueron realizadas en buffer 20 mM Tris-HCl (pH7.5), 200 mM NaCl o 20 mM Formiato de sodio (pH 3.0), 200 mM NaCl a 20 ℃.

En su conjunto, estos resultados apoyan la observación acerca de la naturaleza *pre molten globule* del fragmento P_C, donde la neutralización de las cargas lleva a que la estructura

extendida de P_C adopte reversiblemente un plegamiento globular con un elevado contenido de α-hélices.

II.3.2- Estabilización de estructura α-hélice por TFE.

Para estudiar la estabilización de las regiones con tendencia estructural α-hélice dentro de P utilizamos el co-solvente 2,2,2-trifluoroetanol (TFE), conocido por estabilizar α-hélices pre-formadas. Se realizaron titulaciones en un rango entre 0 y 50 % v/v de TFE a 25 °C. Se escogió trabajar a una temperatura 10 °C mayor a la T_m de P_C para desplazar el equilibrio hacia una estructura más desordenada y evaluar la estabilización de hélice.

Todos los fragmentos de P presentaron la capacidad de estabilizar estructura α-hélice en presencia de concentraciones crecientes de TFE, lo cual es demostrado por el corrimiento del mínimo hacia 208 nm y el aumento en valor absoluto de la elipticidad de la banda a 220 nm (Figura II.22).

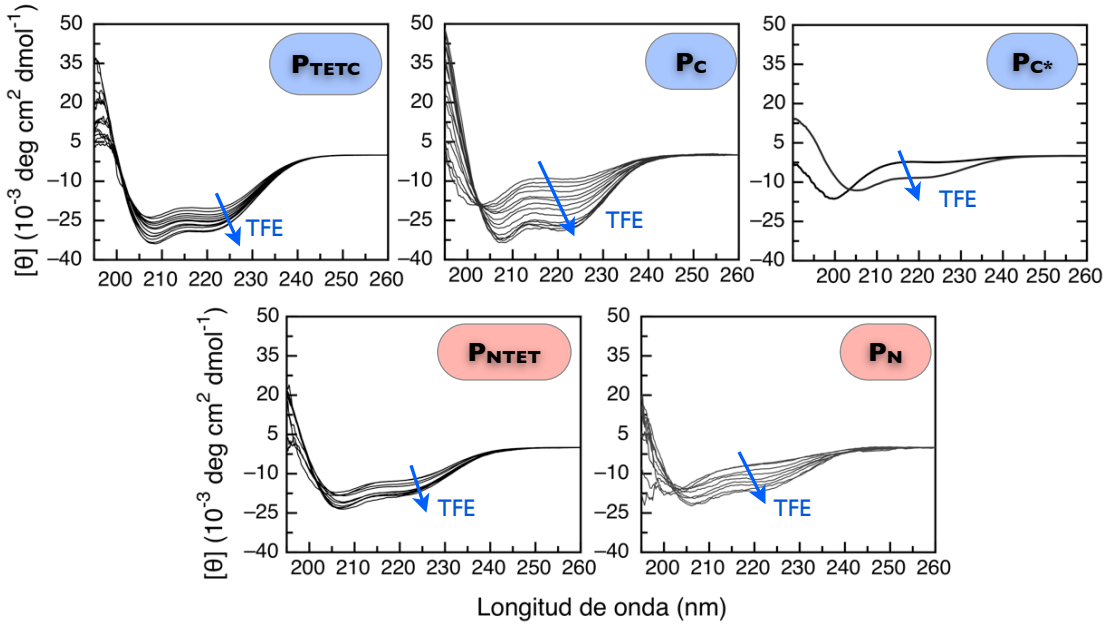


Figura II.22. Análisis de las tendencias de los distintos fragmentons de P a estabilizar α-hélice. Experimentos de titulaciones con TFE seguidos por CD. La muestra fue incubada con TFE 5 minutos previo a medir el espectro de CD. La flecha muestra el sentido de la reacción. Las medidas fueron realizadas en buffer 20 mM Tris-HCl (pH7.5), 50 mM NaCl a 25 ℃.

Las titulaciones con TFE para P_{TETC} y P_C mostraron una marcada estabilización de estructura α-hélice (Figura II.22), en concordancia con la estructura observada en el domino C terminal de P. Por otra parte, ambas transiciones presentan un punto isodicroico alrededor de 203 nm, indicativo de una transición de dos estados. Por otro lado, el fragmento P_{C*}, correspondiente a los últimos 44 residuos del módulo C terminal, en ausencia de TFE no posee ningún componente de estructura secundaria. Sin embargo, en presencia de 50% TFE puede

estabilizar α-hélice (Figura II.22), sugiriendo la presencia de una α-hélice local y fluctuante dentro de dicha región.

Por otro lado, las titulaciones con TFE de las especies P_{NTET} y $P_{N,}$ si bien presentan estabilización de elementos de estructura tipo α-hélice, lo hacen en menor medida y no mostraron un punto isodicroico (Figura II.22), sugiriendo la existencia de al menos dos eventos de nucleación en la estabilización de los segmentos con estructura α-hélice dentro del módulo N terminal.

II.3.3- Integración de las tendencias conformacionales de los módulos de P.

Con el fin de comparar e integrar los resultados de los diferentes abordajes sobre las tendencias conformacionales presentes en cada módulo de P, analizámos el número de residuos estabilizados en α-hélice para las distintas condiciones fisico-químicas estudiadas (por ejemplo a 50% TFE *vs* 0% TFE; pH 3.0 *vs* pH 7.5; 5 °C *vs* 60 °C). Los números de residuos en hélice fueron calculados utilizando los valores de la elipticidad molar a 222 nm en cada condición, basados en estimaciones patrón de literatura [30] (Ver materiales y métodos) (Tabla apéndice A6).

Para la determinación del número de residuos estabilizados en la transición térmica se utilizó como límite superior 60 °C de modo de evitar contribuciones provenientes del desplegado de P_{TET}, el cual ocurre a temperaturas superiores a 65 °C (Figura II.8). Estimamos aproximadamente 59 residuos en conformación α-hélice a 60 °C para P_{TETC}, P_{NTET} y P, los cuales concuerdan con los 52 residuos predichos en conformación hélice según el modelo tridimensional descripto para la región central de P [86].

En primer lugar, observamos que el contenido de estructura α-hélice estabilizada en P_{TETC} a baja temperatura y 50% TFE presenta valores similares de 36 y 33 residuos, respectivamente (Figura II.23, segundo grupo de barras). Por otro lado, el módulo P_C en 50% TFE presenta la estabilización de 43 residuos, un valor mayor al registrado para P_{TETC} (33 residuos) (Figura II. 23, segundo y tercer grupo de barras). Esta diferencia de 10 residuos puede ser explicada por el hecho de que el TFE esta estabilizando residuos en conformación α-hélice que ya están estructurados en el contexto del domino de tetramerización, pero no preformados en el módulo aislado P_C.

Por otra parte, al evaluar la estabilización de estructura hélice a baja temperatura, podemos observar que mientras P_C estabiliza 10 residuos, P_{TETC} esta estabilizando 36. Esta diferencia indica que se requiere toda la información del dominio C terminal para estabilizar dicha estructura. Este resultado también es apoyado por la menor estabilidad térmica presente en P_C y por el hecho de que P_{TETC} es capaz de unir ANS.

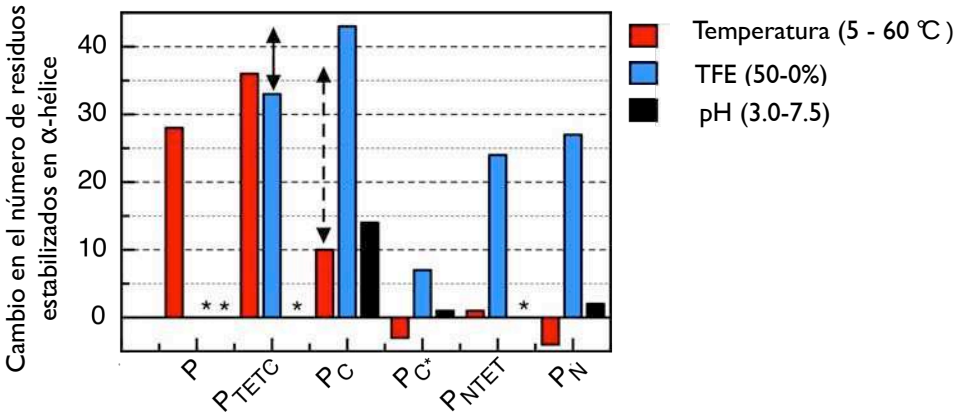


Figura II.23. Análisis integrador del número de residuos estabilizados en α-hélice en los distintos tratamientos. Número de residuos estabilizados por temperatura entre 5 y 60 ℃ (barras rojas), por TFE entre 50% y 0%, (barras azules), o entre pH 3.0 y pH 7.5 (barras negras).

Dentro de P_C, el efecto de reducción del pH (14 residuos estabilizados) es similar en magnitud al efecto estabilizador de la temperatura (10 residuos estabilizados), indicando que el TFE es capaz de estabilizar una mayor proporción de residuos en α-hélice dentro de este módulo, imitando el contexto de la proteína completa (Figura II.23, tercer grupo de barras).

Por otra parte, P_{C*} es desordenado pero con la capacidad de estabilizar estructura PII (Figura II. 9-C), esta tendencia se ve reflejada en los valores negativos observados en la Figura II.23. En presencia de TFE se observó que presenta la capacidad de estabilizar 7 residuos en α-hélice que representa el 16% de los residuos de este fragmento de 44 residuos.

Finalmente, se observó la capacidad de estabilizar α-hélice dentro del módulo N terminal en presencia de TFE, con valores idénticos de números de residuos estabilizados en P_{NTET} y P_N (24 y 27, respectivamente) (Figura II.23, quinto y sexto grupo de barras), confirmando que la hélice transitoria estabilizada en el módulo N terminal de P es independiente del dominio de tetramerización.

DISCUSION

Sobre la base de los resultados obtenidos en el presente capítulo, proponemos el siguiente modelo de equilibrios conformacionales adoptados por P de RSV en respuesta a distintos cambios en el entorno (Figura II.24):

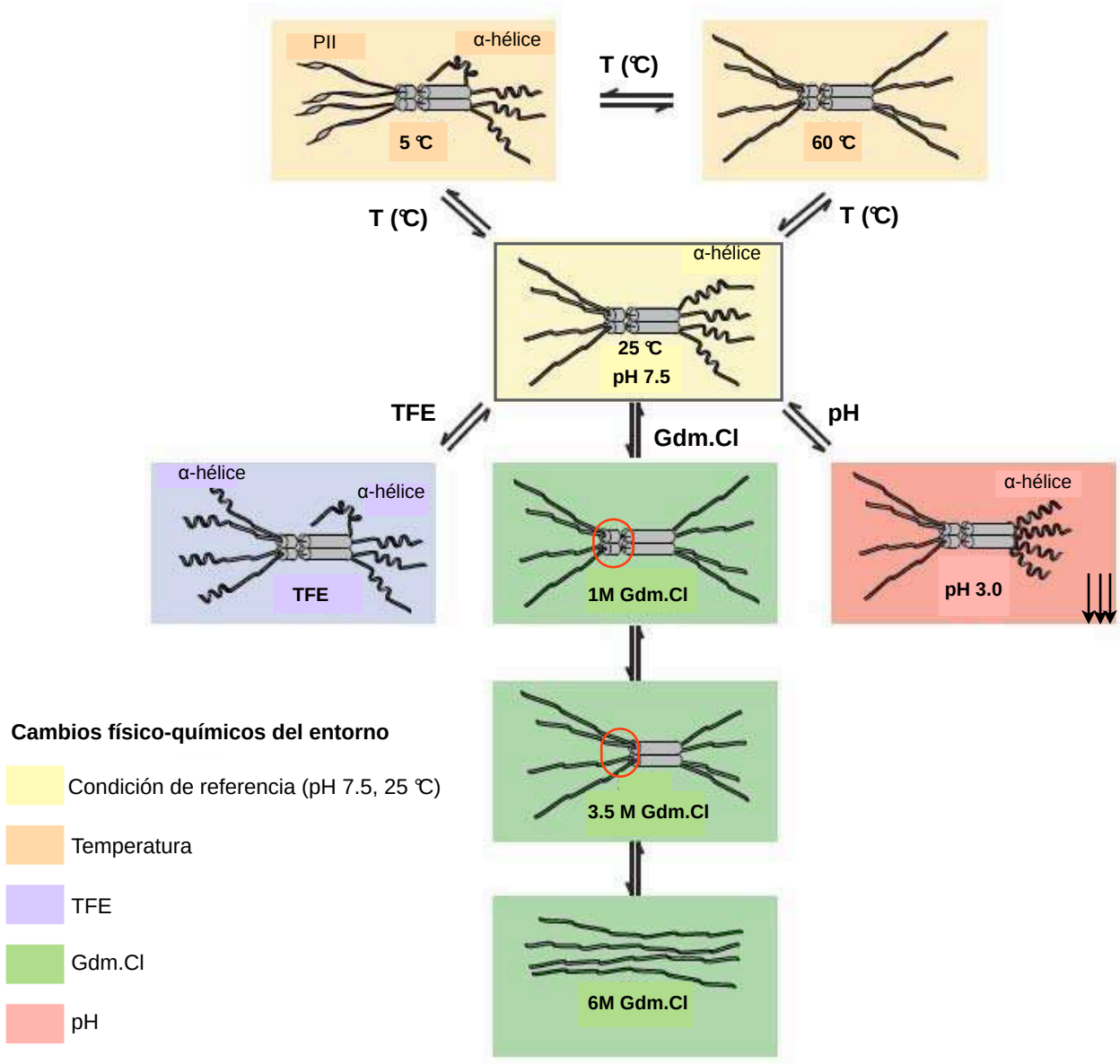


Figura II.24.-Modelo de las conformaciones adoptadas por P de RSV *in vitro*. Recuadrada en gris se muestra la condición de referencia, la cual corresponde a la proteína en condición de buffer pH 7.5 y a 25 ℃. En distintos colores se indican los cambios de entorno a los que se expuso la proteína. El círculo rojo muestra la pérdida de P_Z durante el tratamiento con Gdm Cl.

La fosfoproteína P de RSV es intrínsecamente desordenada.

La fosfoproteína P es un cofactor esencial de la polimerasa viral en todos los miembros de la familia *Paramyxoviridae* [78], y constituye el andamiaje sobre el cual se ensamblan los distintos

componentes del complejo replicativo del virus. Se especula que la misma está involucrada en la regulación de la transición entre la replicación y la transcripción del genoma viral [78]. Se demostró para el virus de la parotiditis humana (MuV), que esta proteína puede desensamblar localmente el complejo NC-ARN para favorecer el acceso de la polimerasa viral al ARN [136]. Por lo tanto, es de esperar que dicha proteína sea flexible y modulable por cambios en el entorno, ya sea por variaciones de temperatura, polaridad del solvente, carga neta por fosforilaciones, etc.

En el presente capítulo demostramos utilizando técnicas bioquímicas y biofísicas y mediante una disección por dominios, que la proteína P de RSV es una proteína intrínsecamente desordenada. Encontramos que la misma posee un gradiente de estabilidad conformacional y de transiciones orden-desorden, que van desde el dominio de tetramerización *coiled-coil* altamente estable, el fragmento P_Z que es resistente a desnaturalización térmica pero se pierde a concentraciones superiores a 1 M Gdm.Cl, el domino C-terminal con características IDP *pre-molten globule y* el módulo N terminal con características de IDP *random-coil* (Figura II.24).

El dominio N-terminal es mayormente desordenado, aunque puede estabilizar localmente elementos de estructura PII así como también tipo α-hélice en presencia de TFE (Figura II.24). Por otra parte, encontramos que el dominio C-terminal presenta características de IDP tipo *pre molten globule*. El mismo posee un elemento de estructura marginalmente estable, cuya estabilidad es modulada por pequeñas variaciones de temperatura, polaridad del entorno o pH (Figura II.24). Hipotetizamos que dicho elemento podría estar actuando como un elemento de estructura preformada α-MORF, potencialmente involucrado en la interacción con los distintos componentes del complejo polimerasa.

Hasta ahora, únicamente elementos estructurales con características α-hélice de diferente estabilidad se han descripto dentro de P de RSV y para los otros miembros de la familia *Paramyxoviridae* en general [137-139]. Proponemos que este tipo de estructuras α-hélice dinámicas fueron seleccionadas evolutivamente por su capacidad de proporcionar una modulación rápida frente a variaciones de temperatura, modificaciones en el entorno físico-químico, modificaciones post-traduccionales e interacciones proteína-proteína. Hipotetizamos que conformaciones tipo lámina-β, si bien son plausibles, no habrían sido seleccionadas a lo largo de la evolución del virus debido a su naturaleza menos dinámica, más estable y más propensa a agregación.

Conservación dentro de la subfamilia *Pneumovirinae.*

Como se mencionó en la introducción, la subfamilia *Pneumovirinae* esta conformada por los géneros *Pneumovirus* y *Metapneumovirus*. Para evaluar la variabilidad evolutiva de los distintos módulos de P dentro de la subfamilia *Pneumovirinae*, realizamos un alineamiento múltiple de

secuencias (N=15) de miembros representativos de dicha subfamilia (Figura Apéndice A13) y calculamos porcentaje de identidad asociada a cada dominio (P_N, P_{TET}, P_C). La visualización del alineamiento se presenta a continuación en forma de *Logobar* [140] (Figura II.25). Dicha representación es similar al logo de secuencia utilizado en el capitulo anterior [122], pero a diferencia del mismo utiliza barras para cada posición de secuencia en lugar de letras, permitiendo analizar perfiles de gran longitud. La altura total de cada posición indica la conservación de secuencia en esa posición medida en bits. Asimismo, el *Logobar* muestra las regiones de *gaps* dentro de la secuencia como barras blancas [140].

Como se puede observar, dentro de P existen distintos grados de conservación de secuencia, donde las regiones menos conservadas corresponden a los dominios intrínsecamente desordenados N y C-terminal, con 40% y 60% de identidad de secuencia, respectivamente. Globalmente podemos observar que la región correspondiente al dominio N-terminal presenta una menor conservación global y un mayor número de *gaps* en la secuencia. Dentro del dominio N terminal la región más conservada corresponde a los primeros residuos del dominio, donde se describió la presencia de un α-MORF que interactúa con N^0. Por otra parte, la región del dominio C-terminal de P (residuos 161-197) adyacente al dominio de tetramerización presenta una conservación de secuencia del 71% semejante a la observada para domino de tetramerización (68%) y superior a la esperada para la región P_{C*} (residuos 197-241), de un 50% (Figura II.25). Esta región adyacente (161-197) corresponde a los primeros 36 residuos del dominio C-terminal IDP. Notablemente, dicha región conservada podría estabilizar estructura hélice local en la fosfoproteína P del metapneumovirus humano (HMPV).

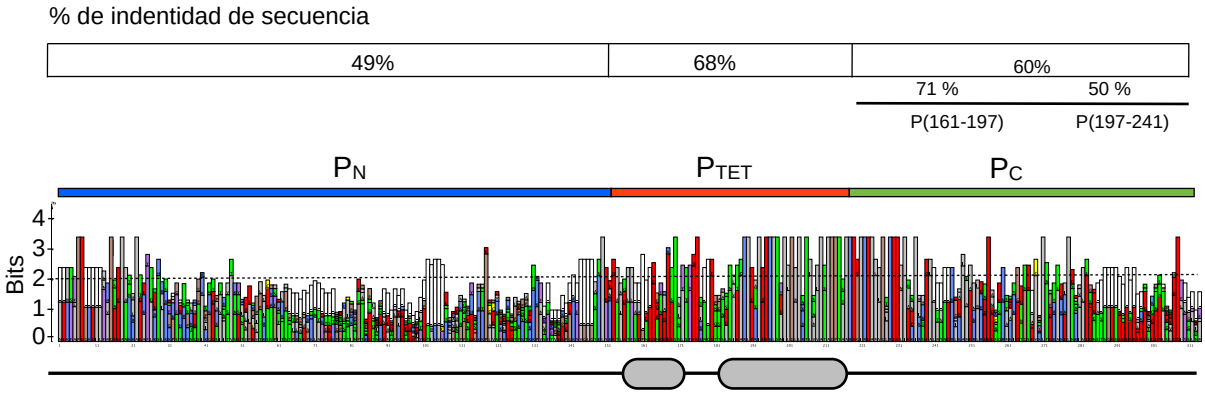


Figura II.25.-Variabilidad evolutiva de P dentro de la subfamilia *Pneumovirinae*. Alineamiento múltiples de 15 secuencias de P, realizados con *ClustalX* y visualizado con *Logobar*. (Género *Pneumovirus*: HRSVA2 (P03421), HRSV B1 (O42062), HRSV Long strain (P12579), ORSV (Q83956), BRSV (P33454), CanineRSV (W0ISD1), DogMPV (F1ACP3), MPV (Q5MKM7). Género *Metapneumovirus*: HMPV strain Can97 (Q8B9Q8), HMPV (Q6QQI3), HMPV (X4XZY6), AMPV_Can(Q2Y2M5), AMPV-C (A8Y7P9), AMPV-B (A8Y7P8), AMPV-D (A0A077SG92). En la parte superior se detalla el % de identidad de secuencia determinado para cada dominio de P calculado a partir de las matrices de identidad de secuencia obtenidas con la herramienta MUSCLE del EMBL. La linea indica el punto de corte de 2 bits.

En un trabajo reciente [141], se propuso un modelo que describe un posible conjunto de conformaciones que puede adoptar la región central de la fosfoproteína P del HMPV

perteneciente al género *Metapneumovirus*. Dicho modelo fue realizado a partir de un modelado *ab-initio* validado por experimentos de SAXS y de dinámica molecular computacional [141]. El mismo describe las posibles conformaciones que puede adoptar dicha región de P de HMPV (Figura II.26-A). Los autores determinaron la presencia de α-hélices transitorias (residuos 195-237) adoptada por los residuos pertenecientes a la secuencia inmediata al dominio de tetramerización (Figura II.26-A, flechas rojas). Además, proponen que las mismas podrían constituir elementos de reconocimiento molecular tipo α-MORFs [141].

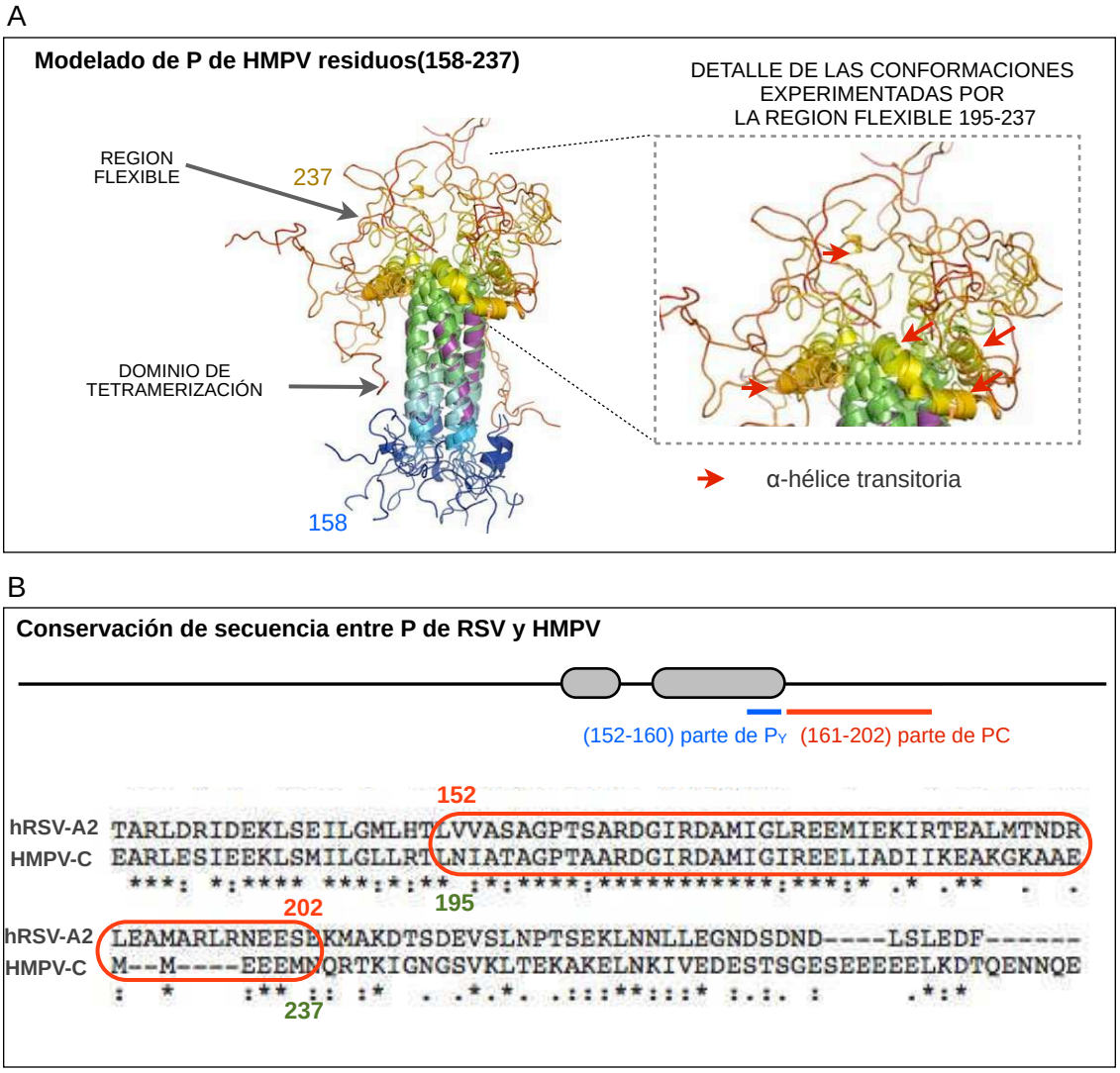


Figura II.26.Características estructurales en común con P de metaneumovirus humano A-Modelo de la región 158-237 de la fosfoproteína P de HMPV. Se detalla la región 195-237 que forma parte del dominio C terminal de P de HMPV, donde las flechas rojas indican la presencia de hélices locales transitorias. Figura adaptada de [141]. B-Conservación de secuencia entre P de RSV y P de HMPV. El recuadro rojo indica el alineamiento de la región flexible de HMPV detallada en A. En el panel superior se esquematiza la posición en P de RSV de la región flexible modelada para HMPV.

Notablemente, la región 195-237 de P de HMPV coincide con la región comprendida entre los residuos 152 a 202 de P de RSV (Figura II.26-B, recuadro rojo), que contiene a los últimos 9 residuos del dominio de tetramerización y 40 residuos del módulo C terminal. En el presente capitulo hemos demostrado por diversos abordajes experimentales y por diversas técnicas

bioquímicas y biofísicas la existencia de un elemento marginalmente estable contenido en el dominio C-terminal de P. Analizando el conjunto de nuestros resultados junto con la conservación de secuencia de esta región y el modelo propuesto para P de HMPV [141], podemos sugerir que la hélice local identificada a partir de datos de SAXS (Figura II.26-A, flecha roja) podría coincidir con el elemento marginalmente estable identificado en el dominio C-terminal de P. Por otra parte, la cercanía con el dominio de tetramerización permitiría explicar, en parte, la estabilización de dicho elemento estructural observada por el contexto tetramérico. Sin embargo, para poder confirmar dicha hipótesis se requiere información a nivel atómico residual solo posible por experimentos de RMN, que se podrán realizar en experimentos a futuro.

P de RSV en el contexto de la subfamilia *Paramyxovirinae*.

Las fosfoproteínas P de la familia *Paramixoviridae* presentan una logitud variable (750 a 241 aminoácidos) (Figura 13, introducción), y baja identidad de secuencia. Sin embargo, poseen una organización modular conservada, son extendidas y poseen un dominio de oligomerización central. Los módulos N-terminal de los virus Nipha (NiV), Hendra (HeV) y Sarampión (MeV) pertenecientes a la subfamilia Paramyxovirinae, han sido identificados como IDPs por distintas técnicas biofísicas [83, 139]. Nuestros resultados acerca de la naturaleza IDP tipo *random-coil* del dominio N-terminal coinciden con estas observaciones y apuntan a una conservación de la conformación extendida de este módulo. En particular, los primeros 50 residuos de las fosfoproteínas P de los virus pertenecientes a la subfamilia *Paramyxovirinae* han sido predichos como α-MORF implicados en la interacción con la proteína de la nucleocápside en su forma monomérica (N^0) [142]. Esta observación ha sido demostrada para el virus Nipha (NiV), donde cristalizaron los primeros 50 residuos de P unido a N^0 y se confirmó que los mismos adoptan conformación tipo α-hélice al interactuar con N^0 [84]. Para RSV, en particular, se identificó recientemente al sitio comprendido entre los residuos 1-29 como el sitio de interacción con N^0, y se propuso que adoptaría conformación tipo α-hélice en complejo con N^0 [96]. Nuestros resultados acerca de la estabilización de una hélice transitoria de aproximadamente 27 residuos dentro del dominio N terminal mediante el agregado de TFE apoyan esta hipótesis.

Por otra parte, como se describió en la introducción de esta tesis, los módulos C terminales de las fosfoproteínas P de las dos subfamilias difieren en su topología (Figura 13, introducción). En particular, las fosfoproteínas P de los virus subfamilia *Paramyxovirinae* contienen regiones globulares intercaladas con regiones flexibles [83]. En general poseen un dominio globular, llamado dominio XD que interactúa con la nucleocápside NC y cuya estructura fue resuelta para los virus MeV, NiV, Hendra y virus Sendai [138, 143, 144]. En particular, el dominio XD del virus de la parotiditis humana (MuV) carece de estructura tridimensional compacta en solución, y se comporta como un *molten globule* que es estabilizado en presencia de co-solventes [29].

Es interesante resaltar dicho dominio *molten globule* de MuV contacta con una región de la NC viral diferente a la descripta para los otros dominios XD de los otros miembros de la subfamilia *Paramyxoviridae* [29]. Nuestros resultados acerca de las características tipo *pre-molten globule* descriptas para el dominio C-terminal de P de RSV constituye un caso intermedio entre una conformación ordenada y desordenada.

Todas las interacciones descriptas entre fosfoproteínas P y nucleocápsides (P-NC) mantienen desorden en el estado unido [99, 145], conformando los llamados complejos *Fuzzy*. Esta conservación de la flexibilidad en el estado unido sugiere una ventaja adaptativa positiva en la función asociada a dicha interacción.

Correlaciones funcionales de las transiciones estructurales estudiadas.

Se ha descripto que la mutación de dos residuos altamente conservados en el dominio C-terminal de P (G172 y E176) producen defectos en la replicación viral *in vitro* e *in vivo* a 37 °C (Figura II.27-A). Sin embargo, dichos mutantes pueden replicar como la especie *wild type* a 33 °C [102]. Un ejemplo de esto se observa en la Figura II.27-B donde se muestra que los virus que contienen dichas mutaciones tienen afectada su capacidad de formar placas de lisis a temperaturas fisiológicas [102]. Dichas mutaciones provocan que cambios de solo 4°C en la temperatura del medio produzcan un efecto deletéreo sobre la capacidad de replicación del virus [102].

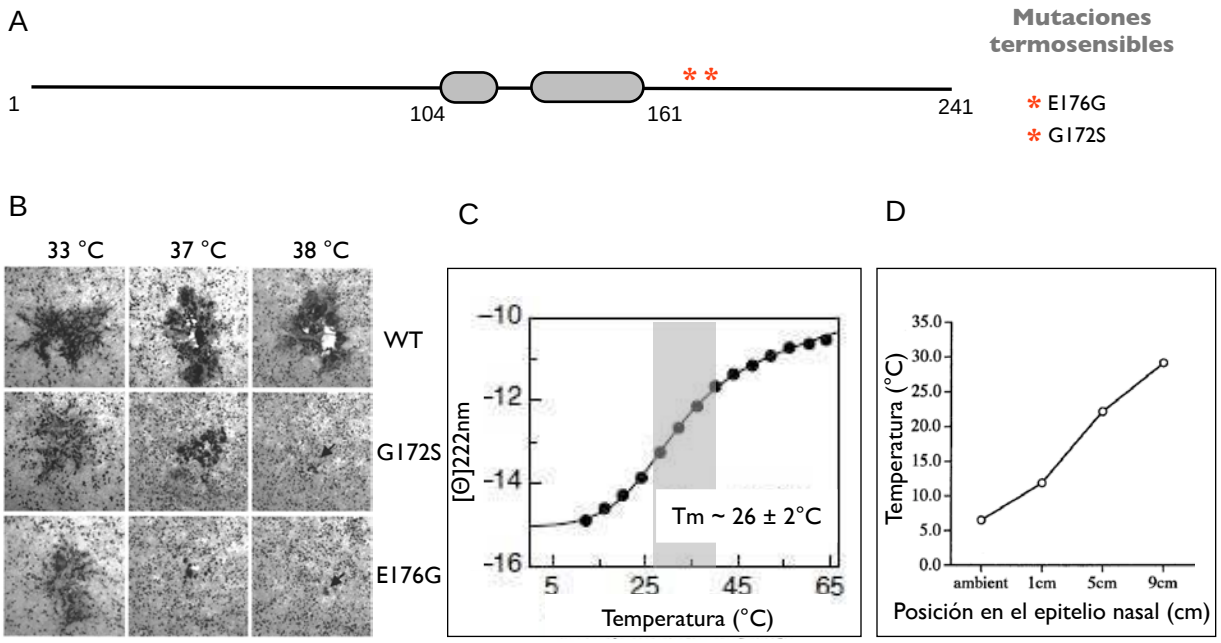


Figura II.27.-Hipótesis del rol de P como proteína termosensora. A-Esquema de la fosfoproteína P de RSV indicando la posición de las mutaciones termosensibles. B-Ensayo de formación de placas realizado en células Hep-2 infectadas con virus RSV WT, y con los virus mutantes G172S y E176G, incubados a distintas temperaturas. Las placas fueron por inmunotinciones con anticuerpos anti-RSV (Adaptado de [102]). C-Transición estructural de P con la temperatura, en gris se marca el rango de temperaturas de 33 a 37 °C. D- Gradiente de temperatura del epitelio nasal del tracto respiratorio superior (Adaptado de [167]).

En el presente capítulo demostramos que P posee un elemento de estructura marginalmente estable, el cual es modulado por pequeñas variaciones de temperaturas (Figura II.27-C). Dichos cambios estructurales en un rango fisiológico de temperatura (Figura II.27-C, barra gris) son rápidos (<10 seg) y reversibles, afectando el contenido de estructura secundaria del dominio C-terminal. Teniendo en cuenta nuestros resultados, sumado a la posible localización de dicho módulo en la región adyacente al domino de tetramerización (residuos 161-202, que contiene a las mutaciones descriptas) y al efecto reportado de las mutaciones (G172 y E176) sobre la replicación del virus, proponemos la hipótesis de que el dominio C-terminal de P podría funcionar como una región regulatoria termosensora durante la replicación y transcripción de RSV. Una observación a favor de esta hipótesis es la existencia de un gradiente de temperaturas a lo largo del tracto respiratorio superior (Figura II.27-D), lo cual sugiere que el virus, a lo largo de su ciclo infeccioso, se encuentra en entornos con variaciones en el rango de temperatura que produce cambios estructurales en P. Como perspectiva futura del presente trabajo, se podría estudiar el efecto de dichas mutaciones sobre la estabilidad térmica del elemento marginalmente estable del domino C-terminal de P. Así como evaluar el efecto de la temperatura sobre la función del complejo polimerasa.

En resumen, la fosfoproteína P es un cofactor esencial de la polimerasa viral L en todos los virus pertenecientes al orden *Mononegavirales*, y puede ser considerada como la proteína de andamiaje esencial requerida para la correcta función del complejo polimerasa. La oligomerización como característica de estas proteínas se encuentra conservada, siendo capaz de formar estructuras tetraméricas, triméricas o diméricas según el virus. Nuestros resultados indican que la tetramerización de P de RSV impacta en la conformación de la proteína y en la estabilidad del dominio C-terminal de P.

El complejo polimerasa de RSV es un blanco atractivo para el desarrollo de antivirales dirigidos a bloquear la replicación y transcripción del genoma viral. Sin embargo, como se mencionó en la introducción de esta tesis, no hay disponible información estructural de la polimerasa viral L y de la fosfoproteína P. Los resultados de este capítulo proveen la primera información disponible acerca de la naturaleza IDP de esta proteína, demostrando la presencia de dos dominios con distinto grado de desorden intrínseco. Estos resultados arraigan la noción de que P de RSV es una proteína extremadamente dinámica. Consideramos que futuros estudios de las distintas conformaciones dinámicas de P de RSV podrían proveer información fundamental acerca del mecanismo del ensamblado del complejo polimerasa viral, los cuales permitirán encontrar blancos a los cuales dirigir el desarrollo de nuevas drogas contra RSV, y contra virus del orden *Mononegavirales* en general.

CONCLUSION GENERAL

CONCLUSIÓN GENERAL

El desorden intrínseco en proteínas es una característica estructural asociada a facilitar la diversificación funcional de las mismas. Las IDPs se encuentran presentes en los tres dominios de la vida: bacteria, eucaria y archea [49]. En particular, representan aproximadamente un tercio de las proteínas del proteoma humano [46] y están asociadas a múltiples funciones celulares, interviniendo en vías de señalización, regulación de la homeostásis celular, etc. Por su rol en la regulación de procesos claves a nivel celular, muchas enfermedades como la diabetes, distintos tipos de cáncer, enfermedades cardiovasculares, neurodegenerativas y enfermedades infecciosas están asociadas a las mismas.

Los virus poseen genomas compactos que codifican para muy pocos productos proteícos. Sin embargo, son capaces de infectar una gran variedad de tejidos y secuestrar la maquinaria de la célula para llevar a cabo su ciclode vida [55]. Por lo tanto, una de las preguntas que motivó la realización de este trabajo de tesis es: ¿Por qué los pocos productos proteicos de los virus pueden interactuar de manera específica con tantos blancos celulares?

En la presente tesis buscamos dar respuesta a este interrogante a través del estudio del dominio E7N de la proteína E7 de HPV y de la fosfoproteína P de RSV. Los resultados generales de esta tesis serán discutidos en el marco de los siguientes conceptos:

(i) Plasticidad estructural en IDPs.

(ii) Papel biológico de los elementos de estructura secundaria local.

(iii) Tipos de desorden intínseco en proteínas.

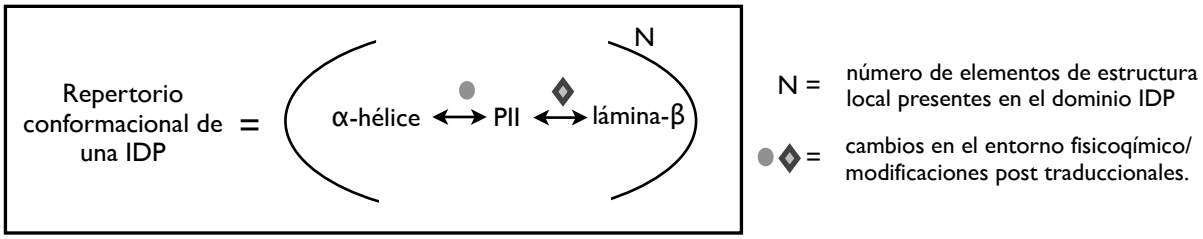
(iv) IDPs y virus.

(i) Plasticidad estuctural en IDPs.

Las IDPs son proteínas dinámicas y flexibles, las cuales contienen regiones con tendencia a adoptar estructuras secundarias localmente que pueden experimentar transiciones del tipo desorden-orden en respuesta a cambios en el entorno físico-químico y biológico en el que se encuentran. En el primer capítulo de esta tesis demostramos que dentro del dominio E7N existen tres regiones definidas con la capacidad de estabilizar localmente elementos de estructura secundaria transitorios frente a cambios en pH, TFE, SDS y temperatura. Por otro lado, en el segundo capítulo demostramos la naturaleza IDP de la proteína P de RSV, y definimos que dicha proteína posee dos dominios N- y C-terminal con distinto grado de desorden intrínseco, que pueden experimentar transiciones desorden-orden, siendo estos equilibrios modificados diferencialmente por cambios en parámetros del entorno como polaridad del medio (SDS, TFE), temperatura y pH.

En ambos modelos estudiados observamos una característica en común: la independencia de los cambios estructurales de distintas regiones frente a una misma modificación en el entorno. Por ejemplo, mostramos que la hélice-I estabilizada en TFE (que se localiza dentro de la región CR1 de E7N) no presenta cambios estructurales inducidos por pH. Sin embargo, la hélice-II estabilizada en TFE, así como la hélice PII estabilizada a bajas temperaturas, son moduladas por cambios en el pH. Asimismo, mientras concentraciones sub-micelares de SDS estabilizan la hélice-I, este no afecta la estabilización de la hélice-II. Por otro lado, demostramos que los módulos IDPs de la fosfoproteína P de RSV no se comportan de la misma manera al ser perturbados a bajas temperaturas o bajo pH. Demostramos que bajas temperaturas estabilizan el contenido de estructura PII en el dominio N-terminal de P, y sin embargo, estabiliza α-hélice en el dominio C-terminal. Además, observamos que la reducción del pH del medio lleva a la globularización del dominio C-terminal con un aumento significativo en el contenido de estructura secundaria α-hélice, observándose una leve compactación del radio hidrodinámico sin efecto sobre la estructura secundaria para el dominio N-terminal. Estas observaciones demuestran la independencia en la "respuesta estructural" presente dentro de una IDP frente a un determinado cambio en el entorno físico-químico.

Estos dos ejemplos nos llevan a preguntarnos: ¿Qué consecuencias funcionales conlleva esta estabilidad diferencial de regiones con tendencia conformacional local? Hipotetizamos que la independencia entre dichas regiones podría aumentar la capacidad combinatoria de regulación de funciones entre sub-regiones o dominios de una misma IDP. De esta manera, la adición de una región independiente que pueda experimentar transiciones del tipo desorden-orden aumentaría exponencialmente el repertorio conformacional de una proteína frente a cambios en el entorno. Esta propiedad permitiría explicar la diversificación de funciones a partir de una misma secuencia primaria, dada por la modulación independiente de cada elemento estructural dentro de una misma IDP (Esquema I).



Esquema I. Repertorio conformacional de una IDP.

(ii) Papel biológico de los elementos de estructura secundaria local.

A lo largo de esta tesis mostramos ejemplos apoyando el concepto de que las regiones o dominios IDPs no son simplemente regiones sin estructura o *random coil*, sino que poseen propiedades características que aportan a su funcionalidad. Demostramos que la región E7N, a

pesar de presentar un radio hidrodinámico mayormente extendido y estructura secundaria característica de un péptido o proteína en estado desplegado, posee la capacidad de estabilizar localmente y en regiones definidas, ciertos tipos de estructura secundaria. La plasticidad de dichos elementos de estructura local, en nuestro caso: hélice-I, hélice-II y estructura PII, serían esenciales para las distintas interacciones mediadas por los mismos.

En un estudio reciente, se demuestra por primera vez el impacto estructural y la correlación funcional *in vivo* de una mutación que afecta la tendencia conformacional de un elemento de estructura secundaria local en una IDP en eucariotas [123]. En dicho trabajo, los autores muestran la importancia de la plasticidad en solución de una hélice transitoria que media la interacción con MDM2 en la región N-terminal de p53. Una mutación puntual en un residuo de prolina que flanquea a dicha hélice transitoria provoca un aumento en la tendencia α-hélice y en la afinidad por MDM2 en aproximadamente un orden de magnitud [123]. Notablemente, una sola mutación que afecta la estabilidad de un elemento de estructura local, tiene un impacto en la función de p53 [123]. Este trabajo pone en evidencia la importancia de los elementos de estructura secundaria transitorios y locales en la función de una IDP.

(iii) Tipos de desorden intrínseco de proteínas.

Como se ha descripto en la introducción de esta tesis, existen distintos grados o tipos de desorden intrínseco en IDPs. Existen ejemplos de proteínas que son totalmente desordenadas, o que presentan elementos de estructura secundaria transitorios, o que pueden existir como estados colapsados tipo *molten globule* [146]. En esta tesis utilizamos la clasificación más actualizada de tipos de desorden las cuales permite categorizarlas como: *IDPs tipo "molten globule"*, *IDPs tipo "pre molten globule"* o *IDPs tipo "random-coil"* [16]. Sin embargo, como las IDPs son dinámicas, esta categorización no representa más que distintos tipos de equilibrios entre transiciones orden y desorden.

Los dominios IDPs estudiados en esta tesis representan ejemplos de distintos tipos de desorden. Por un lado, determinamos que el dominio E7N presenta propiedades hidrodinámicas extendidas y bajo contenido de estructura secundaria en ausencia de co-solventes a pH neutro, sin embargo, presenta la capacidad de estabilizar elementos de estructura secundaria local transitoria. Por sus características el mismo es un ejemplo de IDPs tipo *random-coil*.

Por otro lado, demostramos que P de RSV contiene dos regiones IDPs con distinto tipo de desorden intrínseco. En primer lugar, el dominio N-terminal de P es una IDP tipo *random-coil* la cual no posee estructura secundaria o terciaria consolidada en solución, pero puede estabilizar regiones de estructura α-hélice en presencia de TFE. Dicho dominio contiene cinco de los once sitios de fosforilación descriptos para esta proteína, los cuales podrían impactar en la

modulación de las posibles estructuras que pueda estabilizar el mismo. Por otra parte, el dominio C-terminal de P, presenta características de IDP tipo *pre molten globule*, con módulos de estructura secundaria estables los cuales pueden estabilizar contactos terciarios que favorecen la interacción con ANS. Sin embargo, si bien presenta contactos terciarios los mismo son marginalmente estables, pués concentraciones menores a 1M Gdm.Cl llevan a la pérdida de dicha estructura.

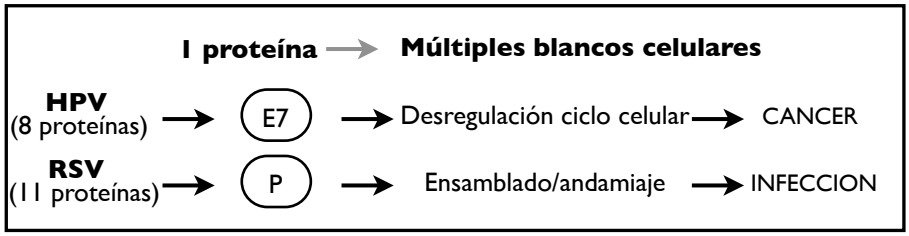
La fosfoproteína P de RSV representa uno de los pocos ejemplos descriptos hasta el momento que contengan dos tipos de regiones IDPs independientes y con distintos grado de desorden intrínseco en su secuencia. Como se mencionó anteriormente, esta proteína cumple una función central en el ensamblado del complejo polimerasa, además de regular la transición entre replicación y transcripción del genoma viral [78]. Por lo tanto, los distintos tipos de desorden y las potenciales transiciones desorden-orden experimentada por esta proteína (Figura II.24) serían fundamentales para su función como proteína de andamiaje.

(iv) IDPs y virus.

Los virus, principalmente los patógenos humanos, presentan genomas compactos, con genes solapados y codifican para pocos productos proteícos [52]. A lo largo de la evolución de los mismos se han seleccionado características que les permitieron optimizar sus ciclos infecciosos [52]. Los elevados contenidos relativos de desorden intrínseco presentes en sus proteomas es una característica que les otorga un valor adaptativo positivo frente a la necesidad de colonizar los distintos tipos celulares infectados. La plasticidad conformacional provista por el desorden intrínseco en conjunto con la capacidad de ser modulado por cambios en el entornos bioquímico, así como por modificaciones post-traduccionales (que pueden funcionar como señales de encendido o apagado de procesos), les otorga una ventaja ideal para poder llevar a cabo sus ciclos de vida en distintos huéspedes, tipos celulares, etc .

A lo largo de esta tesis estudiamos la presencia y el rol del desorden intrínseco en proteínas virales, a través de la caracterización bioquímica y biofísica de dos proteínas pertenecientes a dos modelos de virus de relevancia médica para nuestro país: la oncoproteína E7 de HPV-16 y el cofactor de la polimerasa viral la fosfoproteína P de RSV. (Esquema II). Los modelos estudiados son distintos respecto del tipo de virus que los genera (HPV genoma de ADN y sin envoltura y RSV genoma a ARN y envuelto), del tipo de patología que generan (cáncer, o infección respiratoria), o del tipo de rol que cumple la IDP en la función biológica del virus. En particular los papillomavirus en general no codifican para ninguna proteína con actividad enzimatica, y por lo tanto para poder cumplir su función biológica requieren de la maquinaria replicativa de la célula que infectan. Por lo tanto, su función principal esta asociada a intervenir con vías de señalización. Por otra parte, el RSV codifica para su complejo de replicación y por lo tanto no están asociados a la desregulación del ciclo celular. Ambos modelos estudiados

presentan dos proteínas que cumplen funciones muy distintas en el ciclo biológico del virus que las codifica, sin embargo, las mismas presentan estrategias similares para interactuar con sus distintos blancos.



Esquema II. IDP y virus. Resumen gráfico de las características funcionales de los dos modelos de IDPs estudiados.

Los resultados presentados en esta tesis resaltan la flexibilidad funcional inherente de las IDPs que permite explicar la plétora de procesos biológicos en los cuales se encuentran involucradas.

Tesista

Lic. María Gabriela Noval

Director de Tesis

Dr. Gonzalo de Prat Gay

MATERIALES Y METODOS

I-BIOLOGIA MOLECULAR.

Nota: salvo que se indique lo contrario, todos los protocolos usando enzimas o *kits* comerciales, fueron realizados utilizando las condiciones y tiempos de incubación recomendados por el proveedor.

Vectores de expresión de E7(1-98) y E7(27-98). Se utilizaron las construcciones pMal(E7) [62] y pMal($E7_{27-98}$), previamente desarrollados en nuestro laboratorio. Brevemente; ambas proteínas fueron subclonadas después del C-terminal de la proteína MBP (proteína de unión a maltosa) entre los sitios de restricción BamHI y Hind III, en un vector pMALc2 (New England Biolabs®). Entre el MBP y las proteínas se encuentra un sitio de corte especifico para la proteasa Trombina (LVPRGS), el cual deja como subproducto del corte los residuos GS antes del primer aminoácido de E7 o E7(27-98). Tales aminoácidos no interfieren con ninguna de las propiedades de las proteínas.

Vectores de expresión de P de RSV. Se trabajó con un vector desarrollado previamente en el laboratorio por el Dr. Sebastián Esperante (vector pRSET-A(P)) a partir de un gen sintético comercial (GenScript®) de la fosfoproteína P de RSV subtipo A2 (Código Uniprot: P03421). El mismo fue diseñado con codones optimizados para su expresión en bacterias y posteriormente fue subclonado entre los sitios BamHI y EcoRI (New England Biolabs®) de un vector pRSET-A modificado *ad hoc* conteniendo el sitio de corte para la proteasa trombina (secuencia: LVPRGS) entre el tag de histidina y el sitio múltiple de clonado (Figura apéndice A1).

Desarrollo del vector pRSET-A(P-sintag). Debido a problemas para eliminar el *tag* de histidina durante las primeras pruebas de purificación de las variantes de P realizadas en esta tesis, y dado que la presencia del *tag* no aporta ninguna ventaja al protocolo de purificación de las mismas, se procedió a eliminar el *tag* del vector. Como se muestra en el mapa de secuencia (Figura apéndice A1), la región que codifica para el *tag de* histidina esta flanqueada por sitios de corte para las enzimas NdeI y BamHI, por lo que se decidió eliminar dicho *tag* por digestión con las mismas. En la figura 1 se muestra un esquema del protocolo implementado. Brevemente; el vector pRSETA-(P) fue digerido con las enzimas BamHI y NdeI (GenScript®). El producto de la digestión fue corrido en geles de agarosa 1% y la banda del vector lineal fue cortada y extraída de los mismos utilizando el *kit* de purificación de bandas en gel (GE-Healthcare®). Dado que ambas enzimas utilizadas para la digestión dejan extremos cohesivos no complementarios, se trato con el fragmento de Klenow de la ADN polimerasa I (New England Biolabs®) para generar extremos romos. Luego, el producto fue ligado con la enzima ADN T4 ligasa (New England Biolabs®) durante toda la noche a temperatura ambiente y posteriormente fue transformado en la cepa de clonado DH5α.

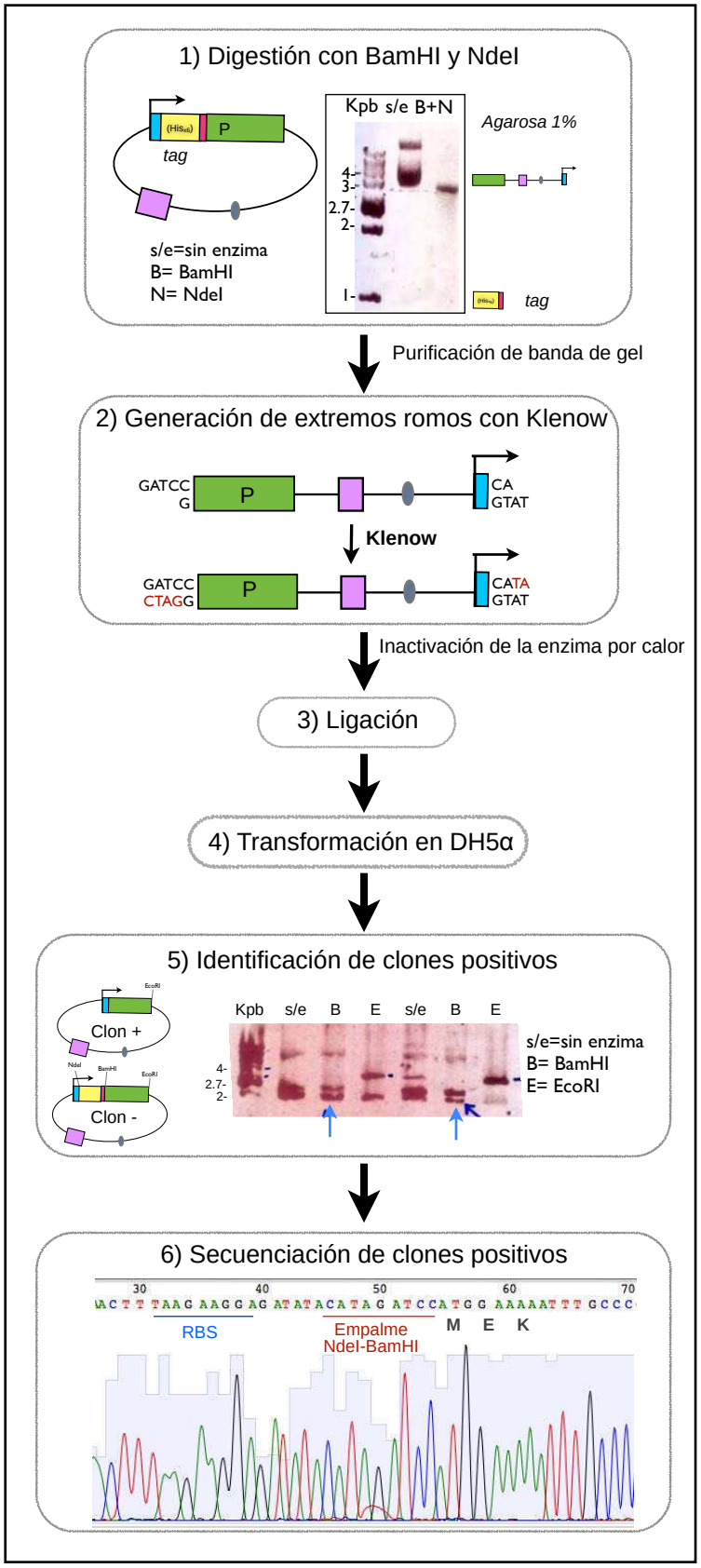


Figura 1. Esquema implementado para la remoción del *tag de histidina.*

Los clones positivos fueron identificados analizando el patrón de digestión (Figura 1). Se consideraron clones positivos a aquellos clones que presentan anulado el sitio de corte para BamHI. Los mismos fueron confirmados por secuenciación. Nota: a pesar de haber eliminado el *tag*, la distancia entre el sitio de unión al ribosoma bacteriano (RBS) y el ATG de P se encuentran a una distancia óptima para que se pueda dar el inicio de la traducción (aproximadamente 14 nucleótidos) (Figura 1).

Desarrollo de los clones PTrombina-I y PTrombina-II por mutagénesis sitio dirigida. Para generar fragmentos conteniendo distintos módulos de P a partir de una misma purificación, decidimos implementar la estrategia de agregar sitios de corte internos para la enzima trombina (LVPRGS). Para ello, utilizando como molde al vector pRSET-A(P-sintag) generado en esta tesis, desarrollamos dos construcciones a las que llamamos pTrombina-I y pTrombina-II. La construcción pTrombina-I fue diseñada para generar los fragmentos correspondientes al módulo N terminal de P (P_N) y al dominio de tetramerización más el módulo C terminal (P_{TETC}). Por otra parte, la construcción pTrombina-II fue diseñada para generar los fragmentos correspondientes al módulo N terminal más el dominio de tetramerización (P_{NTET}) y el módulo C terminal (P_C). En la figura 2 se indican las posiciones y las secuencias de los *primers* utilizados para dar origen a cada construcción.

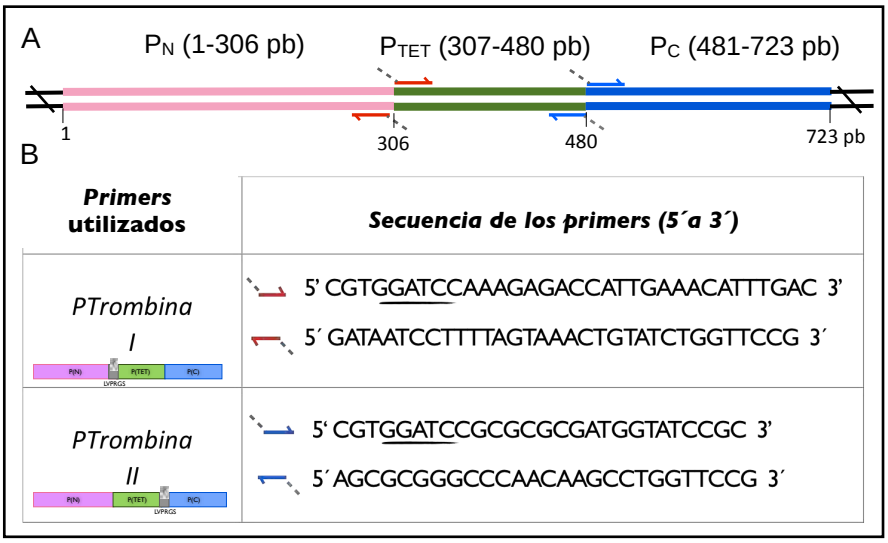


Figura 2. Esquema del diseño de primers para la mutagénesis sitio dirigida. A-Esquema del gen de P indicando las posiciones donde fueron realizadas las inserciones de los sitio de corte para trombina. B-Secuencia de los primer utilizados. Subrayado se indica el sitio de corte para la enzima BamHI. En rosa, verde y azul se indican la regiones correspondientes al módulos P_N, P_{TET} y P_C, respectivamente.

El protocolo general utilizado para la mutagénesis se detalla en la figura 3. Brevemente; los *primers* utilizados para cada reacción fueron fosforilados al 5' por la enzima la polinucleótido kinasa T4 (New England Biolabs®). Luego utilizando como molde al vector pRSET-A(Psintag) se realizaron los ciclos de reacción en cadena de la polimerasa (PCR, por sus siglas en inglés) indicados en la figura. El producto de PCR fue posteriormente tratado con la enzima Dpn-I (New England Biolabs®) para degradar vector molde que no contiene la inserción deseada. El

producto lineal, fue ligado toda la noche a temperatura ambiente y 12 horas más tarde el fue transformado en la cepa de clonado DH5α. La identificación de los clones positivos fue realizada a través del análisis de los perfiles de restricción al cortar con EcoRI y BamHI (Figura 3-B). Se consideraron clones positivos, a aquellos clones que presentan doble banda en el perfil de digestión. Los clones positivos fueron confirmados por secuenciación (Figura 3-B y apéndice A2). Nota: para facilitar la identificación de aquellos clones que incorporaron la mutación, los *primers* fueron diseñados conteniendo un sitio de corte interno para la enzima BamHI (Figura 2.-B, subrayado).

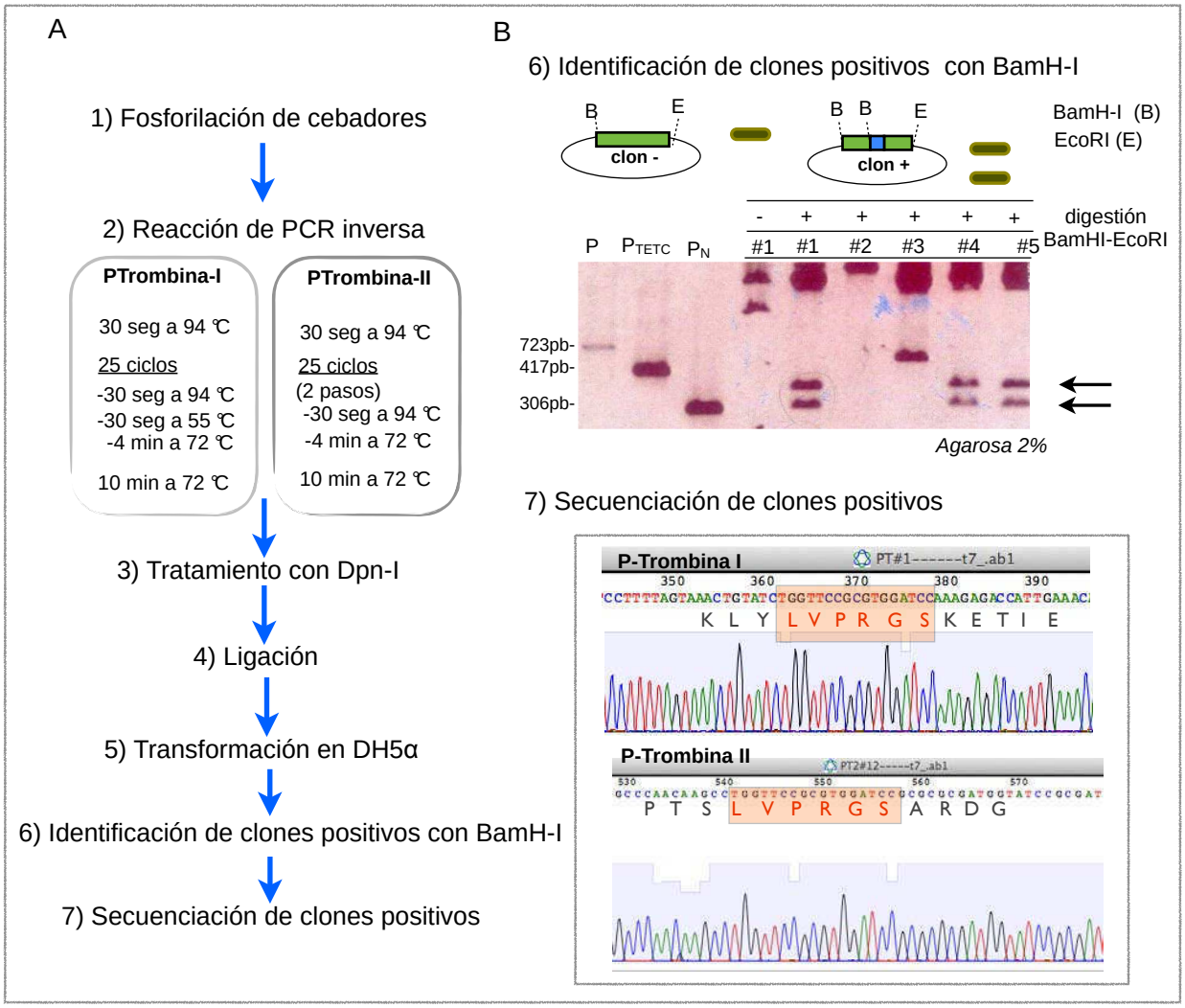


Figura 3.Desarrollo de las variantes P-Trombina I y II. A-Esquema del protocolo de mutagénesis sitio dirigida utilizado. B-Identificación y confirmación de clones positivos.

II-EXPRESIÓN Y PURIFICACIÓN DE PROTEÍNAS.

Expresión y purificación de la proteína E7 y la variante truncada E7(27-98) de HPV-16. *Expresión.* Para la expresión de estas proteínas, se utilizó el protocolo de purificación de la proteína E7 con fusión a MBP desarrollado previamente en el laboratorio [62].

Protocolo de expresión y purificación de P de RSV y de las variantes PTrombinal y II. Las construcciones pRSET-A(P-sintag), pRSET-A(P-sintag-PTI) y pRSET-A(P-sintag-PTII), fueron expresadas y purificadas utilizando como base un protocolo previamente descripto para pRSET-A(P), sobre el cual se realizaron adaptaciones según las particularidades de cada construcción (Figura 4). *Expresión.* Células la cepa de *E.coli* C41(DE_3) fueron transformadas químicamente con ~300 µg de vector. Una colonia fue transferida a 500 ml de medio Terrific Broth con 100 µg/ml de ampicilina (TB-Ampi) suplementado con 0.3% glucosa. El cultivo fue crecido toda la noche a 220 rpm y 37 ºC hasta llegar a saturación. Doce horas post inoculación, el cultivo fue diluido a la mitad con 500 ml de medio TB-Ampi sin glucosa. El litro de cultivo diluido fue dividido de a 500 ml de cultivo por *erlenmayer*, los cuales fueron incubados en agitación por 20 minutos a 37 ºC. Los cultivos fueron inducidos con 1 mM IPTG y las células fueron cosechadas 3 horas post-inducción en centrifuga Sorvall a 4000 rpm por 20 minutos. Los precipitados bacterianos fueron resuspendidos en 100 ml de buffer de lisis (50 mM Tris-HCl pH 8.0, 0.6 M NaCl, 5 mM 2-mercapto etanol, 1mM PMSF, 1 mM EDTA) y fueron guardados a -80 ºC para proseguir con el protocolo de purificación. *Purificación.* Las bacterias fueron sometidas a 2 rondas de congelado/descongelado y posteriormente fueron lisadas por sonicación, y centrifugadas a no menos de 12.000 rpm a 4 ºC durante 20 minutos. Nota: como la proteína posee expresión soluble se continúo trabajando con dichas fracciones. La proteína fue precipitada con 40% $(NH_4)_2SO_4$ por una hora a 4 ºC, y centrifugada a máxima velocidad por 25 minutos. El precipitado fue resuspendido en 30 ml de buffer de diálisis (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 1mM PMSF) y fue dializado toda la noche en cuarto frío contra 1 litro de buffer, realizando un segundo cambio de buffer al día siguiente. La muestra fue centrifugada para eliminar posibles agregados. En este punto del protocolo se tomo ventaja de la característica estabilidad térmica de las IDPs, y por lo tanto el soluble fue sometido a un shock de temperatura a 75 ºC por 20 minutos. Luego la muestra fue transferida a hielo por 10 minutos y ultracentrifugada por 30 minutos a 40.000 rpm. La fracción soluble resultante fue tratada con 1 mg de la enzima Ribonucleasa A (Sigma®) y con 72U de DNAsa-I (Sigma®) durante 4 horas a 37 ºC. La muestra fue dializada toda la noche contra 1 litro de buffer de columna (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 1mM PMSF) (Figura 4). *Intercambio iónico.* Los 30 ml de muestra fueron inyectados en una columna de intercambio aniónico QHyperD (Pall®) en buffer 50 mM Tris-HCl pH 8.0, 50 mM NaCl. Fueron realizados dos lavados consecutivos en 150 mM NaCl (10% B) y otro en 250 mM NaCl (20% B), y posteriormente fue aplicado un gradiente entre 250

mM a 1M NaCl (20% B a 100% B) en 10 volúmenes de columna. Las proteínas P, PTI y PTII eluyeron a una concentración de NaCl entre 300 mM y 500 mM (Figura 5-A). A partir de este punto los protocolos de las variantes PTrombina se separaron (Figura 4).

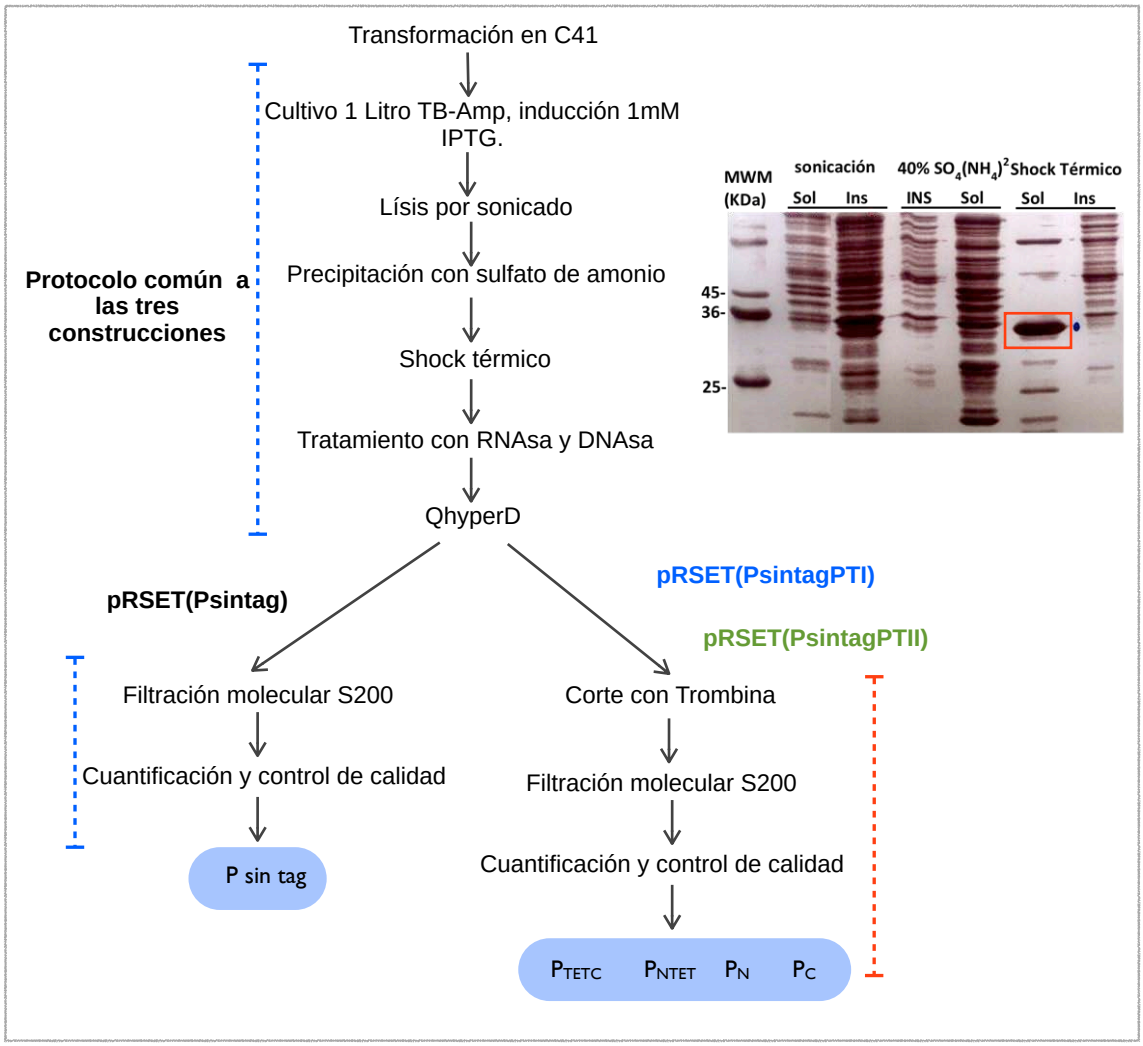


Figura 4. Esquema de purificación desarrollado para Psintag, PTrombina-I y PTrombina-II.

<u>Purificación de P sin tag.</u> La muestra proveniente de la columna QhyperD fue dializada contra buffer (50 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM PMSF) y fue concentrada a 2 ml en concentradores de centrifuga Amicon (Millipore®). *Filtración molecular.* La proteína fue corrida en una columna de tamiz molecular Superdex 200 preparativa (Amersham-Pharmacia®), utilizando un equipo FPLC. *Guardado de la muestra.* La proteína fue dializada en buffer de guardado (50 mM Tris-HCl pH 7.5, 50 mM NaCl) toda la noche a 4 ºC y concentrada a 1 mg/ml. El rendimiento promedio de la purificación fue de 9 mg/L de cultivo. Los stocks concentrados de proteína fueron chequeados por SDS-PAGE 15% para estimar su pureza, corridos en una columna de filtración analítica, confirmados por MALDI-TOF y guardados a -80 ºC (ver apéndice, Figura A4). Nota: esta muestra también fue sometida a pruebas de liofilización y resuspensión, obteniendo un porcentaje de recuperación > 90%. Por su carácter IDP esta

proteína es muy sensible a degradación por cualquier tipo de proteasas presentes en el material de trabajo. Es recomendable trabajar con buffers, agua y todo material líquido autoclavado. Si es posible, trabajar en condiciones de inhibidores de proteasas (1mM PMSF).

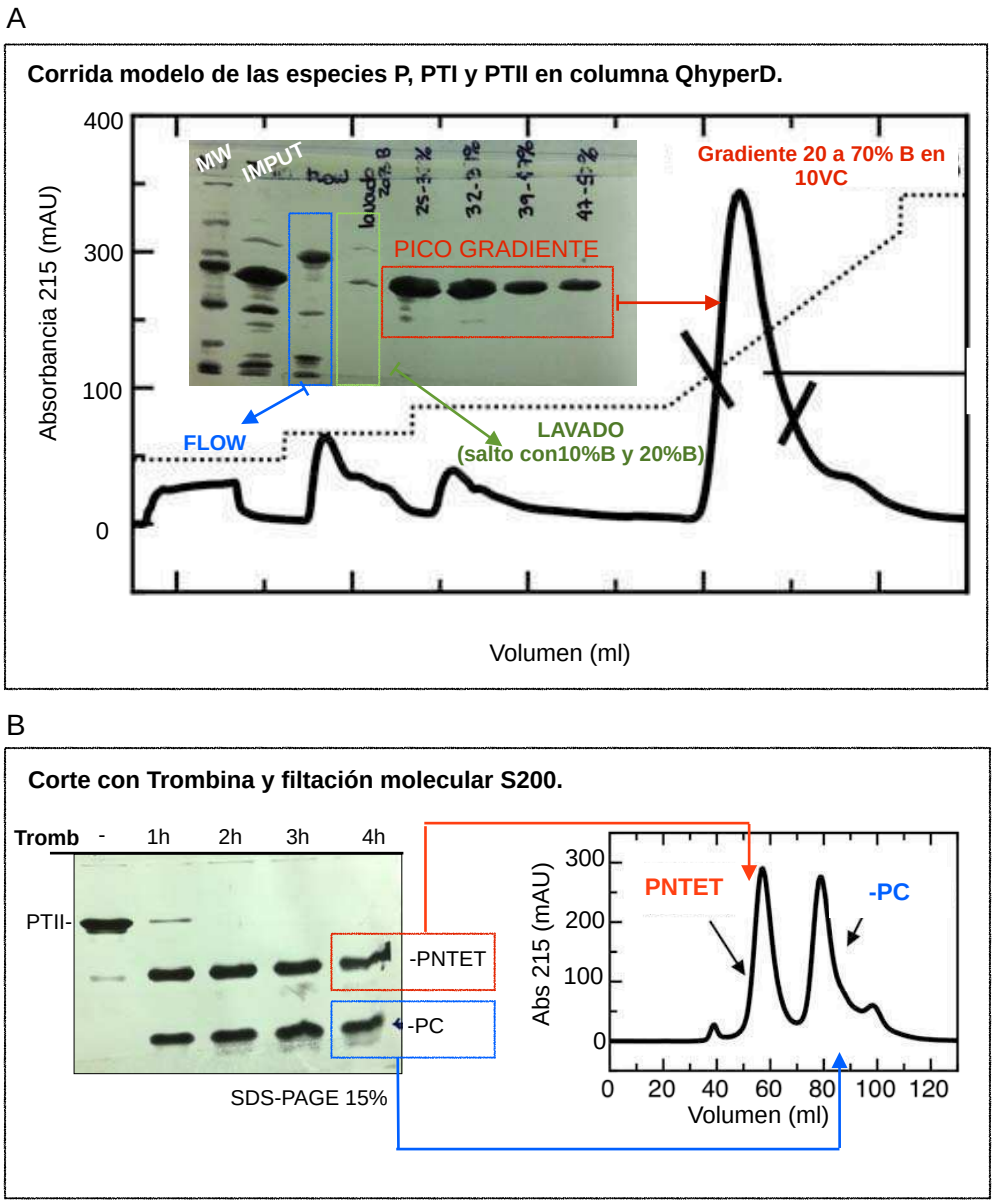


Figura 5. Ejemplo de purificación de PTrombina II. A-Cromatografía de intercambio iónico de PTII. Se muestra el detalle del cromatograma obtenido y el SDS-PAGE de cada una de las fracciones. El recuadro rojo indica las fracciones correspondientes PTII. B-Análisis de la digestión con Trombina por SDS-PAGE y separación de picos por S200. Se indica en el cromatograma la separación de los picos, lo que permite obtener fracciones puras de cada fragmento.

Purificación PTI y PTII: obtención de los distintos fragmentos de P (P_N, P_{TETC}, P_{NTET} y P_C). La muestras de PTI y PTII provenientes del paso de intercambió iónico, fueron dializadas toda la noche contra buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl) y posteriormente concentradas a ~3 mg/ml y digeridas con Trombina (Sigma, 1U/mg) durante 3 horas a 37 ºC. La digestión fue

realizada en buffer 50 mM Tris-HCl pH 7.5, 50 mM NaCl, 2.5 mM $CaCl_2$ y frenada por el agregado de 1 mM PMSF. Nota: a partir de la digestión de PTI y PTII se obtuvieron los fragmentos (P_N + P_{TETC}) y (P_{NTET} + P_C), respectivamente (Figura 5). *Filtración molecular.* Las muestras fueron llevadas, por agregado de NaCl sólido, a una concentración de NaCl de 0.2 M y fueron corridas a un flujo de 2 ml/min en una columna Superdex 200 preparativa (Amersham-Pharmacia®), donde fueron separados los picos correspondientes a cada subfragmento (Figura 5). *Guardado de la muestra.* Los distintos sub-fragmentos fueron dializados en buffer de guardado (50 mM Tris-HCl pH 7.5, 50 mM NaCl) toda la noche a 4 ºC y concentrada a 1 mg/ml. El rendimiento promedio de la purificación por fragmento fue de 3 mg por litro de cultivo. Los stocks concentrados de los distintos sub-fragmentos fueron chequeados por SDS-PAGE 15% para estimar su pureza, corridos en una columna de filtración analítica, confirmados por MALDI-TOF y guardados a -80 ºC (Figura A5-A8). Nota: se observaron menores rendimientos en la producción de proteína expresada a partir del vector pRSET-A(P-sintag-PTI) en comparación con los otros vectores, posiblemente debido a alguna mutación en el vector durante la mutagénesis sitio dirigida. Si bien a los fines de esta tesis la masa obtenida de PTI fue suficiente para los experimentos, es importante considerar hacer pruebas de expresión sobre otros clones del mismo vector para ver si se puede mejorar los rendimientos. Por otra parte, todos los fragmentos fueron sometidos a pruebas de liofilización y resuspensión, obteniendo un porcentaje de recuperación > 90%. Por su carácter IDP esta proteína es muy sensible a degradación por cualquier tipo de proteasas presentes en el material de trabajo. Es recomendable trabajar con *buffers*, agua y todo material líquido autoclavado. Si es posible, trabajar en condiciones de inhibidores de proteasas (1mM PMSF). Muchas veces se ha observado co-purificación con una banda de 60 KDa, en los casos que se observaron esas bandas o presencia de bandas de degradación, las muestras fueron posteriormente re-purificadas utilizando una columna Mono-Q análitica

Obtención de los fragmentos P_Y y P_{C*} por proteólisis triptica. El fragmento P_Y fue obtenido a partir del protocolo de proteólisis limitada de la proteína completa previamente descripto [132]. Brevemente; proteína P fue digerida con tripsina de pancreas bovino (Sigma-Aldrich) 1:100 masa de tripsina/masa de proteína, en 100 mM Tris-HCl (pH 7.5), 50 mM NaCl por 2 horas a 37 ºC. La reacción fue detenida por el agregado de 1 mM PMSF. La reacción fue seguida por SDS PAGE 15%. La muestra fue dializada contra buffer 20 mM Tris-Cl pH 8.0, 200 mM NaCl, y pasado por una columna de filtración molecular Superdex 75 preparativa (Amersham-Pharmacia®). El fragmento P_Y fue concentrado a 1 mg/ml. Por otra parte, el subfragmento de P_C (P_{C*}) fue obtenido a partir de la digestión con tripsina TPCK (Sigma-Aldrich) en 100 mM Tris-Cl pH 8.0, 50 mM NaCl y 50 mM $CaCl_2$ por 45 minutos a 37 ºC, en una relación 7500:1 masa de proteína/masa de tripsina (Figura apéndice A9). La reacción fue detenida por el agregado de 1 mM PMSF, y el producto de digestión fue purificado por cromatografía de fase reversa utilizando una columna C8 (Restek).

Control de calidad y cuantificación de las proteínas utilizadas. La pureza de las distintas proteínas y fragmentos fueron determinadas por SDS-PAGE y perfiles de filtración en tamiz molecular SuperdexS200 analítica. Para todos los experimentos se trabajo con muestras de pureza superior al 95%. Los pesos moleculares de las distintas proteínas fueron verificados por MALDI-TOF (Ver apéndice, Figuras A3-A8). El correcto plegamiento y el estado de oligomerización de cada una de las especies estudiadas fueron evaluados por dicroísmo circular al UV lejano, medidas de dispersión de luz estática (SLS) y mediante corridas de filtración molecular (SEC).

Determinación de la concentración de E7 y de E7(27-98) de HPV-16. La cuantificación de ambas especies fue realizada por el método de Bradford. Se determinó la concentración utilizando una curva de calibración usando como proteína estándar a la seroalbúmina bovina (BSA) (Invitrogen®) en el rango 0-10 µg/ml. La absorbancia de la muestra patrón (800 µl), fue medida luego de agregar 200 µl de reactivo de Bradford concentrado (Biorad®). La curva de calibración se realizó por duplicado y las concentraciones de las muestras incógnitas fueron determinadas por triplicado. Los valores de absorbancia fueron utilizados para estimar la concentración de proteína interpolando en la curva patrón.

Determinación de la concentración de la fosfoproteína P y sus fragmentos. La determinación de la concentración de P y los distintos fragmentos fue puesta a punto a través de un análisis detallado de la mejor condición de cuantificación que permitiera realizar posteriormente una comparación fina entre los distintos fragmentos de P. Debido a que éstos presentan un tamaño variable (entre 80 y 241 residuos), no poseen Trp, presentan un contenido variable de Tyr (0 a 4), poseen distintos puntos isoélectricos (pI 4.34 a 6.58) y presentan distinto contenido de desorden, métodos de cuantificación basados en absorbancia de tirosinas o Bradford no resultaron confiables. Por lo tanto, fue utilizada la técnica de determinación de concentración de enlace amídico por absorbancia a 220 nm. Originariamente esta técnica fue puesta a punto para E7 en nuestro laboratorio y consiste en cuantificar la muestra en solución ácida 1% v/v HCl. Sin embargo, como fue observada agregación a pH ácido para algunos de los fragmentos de P, la técnica fue adaptada a pH neutro. Fueron analizadas diversas curvas patrones de distintas proteínas (BSA, E1A, E7) a pH ácido y neutro. Dichas proteínas estándar presentan un grado de compactación variable (globulares e IDPs) y distinto número de enlaces (606 a 113). Las curvas de calibración fueron realizadas con BSA y E1A abarcando un rango entre 0 y 600 µM de enlaces y por duplicado, sin observar diferencias entre los distintos estándares . Las muestras fueron preparadas por triplicado en solución acuosa a pH neutro e incubadas por 15 minutos antes de realizar la determinación por absorbancia a 220 nm.

III-PÉPTIDOS SINTÉTICOS DE E7.

Los péptidos de E7 fueron obtenidos por síntesis en fase sólida en un servicio de la Universidad de Yale, Estados Unidos. Estos péptidos fueron purificados por HPLC y analizados por MALDI-TOF para confirmar su peso molecular y su grado de pureza. Todos los péptidos conteniendo la Cys 24 fueron guardados en buffer con 20 mM DTT para evitar la formación de dímeros oxidados, un comportamiento muy frecuente en estas especies. La concentración de los distintos péptidos fue realizada espectrofotometricamente utilizando el coeficiente de absorción molar de los péptidos estimados por la herramienta *ProtParam* (http://web.expasy.org/protparam/). Además la cuantificación de los péptidos fue confirmada por la determinación de la absortividad del enlace amídico a 220 nm en 1% v/v HCl utilizando como patrón a la proteína CI2 (Inhibidor de la chymotripsina 2). Se midió la absorbancia de las muestras incógnita por triplicado y la concentración fue estimada por interpolación en la curva de calibración. En la tabla se detallan las secuencias, pesos moleculares y las absortividades molares (ε_{280nm}) a 280 nm de cada uno de los péptidos de E7 utilizados.

Péptido	***Secuencia aminoacídica***	***Peso molecular (Da)***	***ε_{280nm} ($M^{-1}cm^{-1}$)***
E7N (E7(1-40))	MHG**D**TPTLH**E**YML**D**LQP**E**TT**D**LYCY**E**QLN**D**SS**EEEDE**I**D**G	4677.9	4470
CR1 (E7(1-20))	MHG**D**TPTLH**E**YML**D**LQP**E**TT	2229.5	1490
CR2(E7(16-40))	QP**E**TT**D**LYCY**E**QLN**D**SS**EEEDE**I**D**G	2922.9	2980
E7(25-40)	**Y**E**Q**LN**D**SS**EEEDE**I**D**G	1871.8	1490
E7(16-31)	QP**E**TT**D**LYCY**E**QLN**D**S	1919	2980
E7(21-29)	**D**LYCY**E**QLN	1160.2	2980

IV-DIGESTIONES TRIPTICAS Y ESPECTROMETRIA DE MASA

Digestión con tripsina de la fosfoproteína P y sus fragmentos. Para los estudios de proteólisis limitada fue utilizado un stock (1µg/ml) de tripsina de pancreas bovino tratada con TPCK (Sigma-Aldrich®). Se realizaron digestiones analíticas a partir de 10 µg de proteína en buffer 100mM Tris-HCl (pH 8), 50 mM $CaCl_2$ a 37℃. Fueron barridas distintas relaciones m/m (masa de tripsina/masa de proteína) abarcando rangos de 1:1000000 a 1:10 a tiempo fijo de 1 hora, utilizando como control la poteína BSA. Los productos de digestión fueron resueltos en geles SDS-PAGE al 18%. Las condiciones elegidas para el ensayo de digestión de los fragmentos P_{TETC} y P_C, fueron 37℃, 1:7500 m/m (masa de tripsina/masa de proteína) a distintos tiempos a una concentración final de proteína de 10 µM en un volumen final de reacción de 100µl. Esta reacción fue seguida por gel de SDS 18% y por cromatografía líquida

de alta presión HPLC, en columna de fase reversa analítica (C18) (Restek®). Posteriormente, las especies aisladas fueron identificadas por espectroscopía de masa MALDI-TOF.

Análisis de péptidos y proteínas por espectrometría de masa MALDI-TOF. Las muestras conteniendo péptidos o fragmentos de proteínas de hasta 3 KDa fueron analizados utilizando resina de ácido α-ciano- 4-hidroxicinámico (HCCA) (Sigma-Aldrich®). Para las proteínas o fragmentos de mayor tamaño fue utilizada la resina de ácido sinapínico (SA) (Sigma-Aldrich®). Las muestras fueron preparadas mezclando 2µl demuestra de péptido ó proteína en 0.1% TFA con 2µl de solución de resina (60% acetonitrilo/0.1% TFA). Para la calibración del equipo fueron utilizados kits de calibración comercial "Peptide calibration standard I" (Bruker®) ~1 KDa a 3.5 KDa, ó "Protein calibration standard II" (Bruker) ~10 KDa a 70 KDa. Los espectros fueron obtenidos en un equipo MALDI-TOF Microflex (Bruker, Billerica MA). Todas las mediciones fueron realizadas utilizando el programa Flex-Control (Bruker®) en modo reflectrón positivo. Se utilizó el programa Flex-Análisis para el análisis de los espectros y el programa Peptide-Cutter (http://www.expasy.ch/tools/peptidecutter) para obtener los patrones de digestión tríptica teóricos para las diferentes proteínas.

V-GELES SDS PAGE, REACTIVOS Y DROGAS PARA ESPECTROSCOPÍA.

Todos los SDS-PAGE fueron realizados según el método de Laemmli. Previo a la siembra de las muestras en el gel, éstas fueron tratadas con un buffer; 125 mM Tris-HCl, pH 6.8; 4% SDS; 0.02% Azul de Bromofenol; 25% v/v glicerol, 5mM EDTA. Los geles fueron teñidos con una solución de Coomassie Blue.

Todos los reactivos utilizados fueron de calidad espectroscópica: GdmCl (Invitrogen®), el SDS (Bio-Rad®), Tris (USB®), Fosfato (Sigma-Aldrich®), Formiato (Sigma-Aldrich®), NaCl (USB®), ANS (Sigma-Aldrich®), TFE (Sigma-Aldrich®), DTT (Sigma-Aldrich®). Todos los buffers fueron preparados el mismo día del experimento y filtrados con filtro de jeringa de 0.22 µm.

VI-ESTUDIOS CONFORMACIONALES POR DICROÍSMO CIRCULAR.

Las mediciones de dicroísmo circular al UV-lejano (CD) fueron realizadas en un espectropolarímetro JASCO J-810 (Jasco®, Japón) utilizando cubetas de cuarzo 0.1, 0.2, 0.5 y 1 cm de paso óptico. Los espectros fueron medidos en rangos pertenecientes al UV-lejano (entre 195 y 260 nm), empleando una velocidad de barrido de 50 nm/min, un ancho de banda de 4 nm, y un tiempo de respuesta promedio de 8 segundos. La temperatura fue regulada en un rango de ± 0.1 ºC de la temperatura seleccionada, utilizando un dispositivo peltier. Los

espectros de CD resultantes provienen de un promedio de 6 mediciones. Los datos crudos fueron convertidos elipticidad molar media por residuo (MRW) según la siguiente ecuación [30]:

[Ecuación 1]

$$[\theta]_{MWR} = \frac{\text{deg}}{(\#_{bounds} \cdot [c] \cdot 10000 \cdot L)}$$

Donde *deg* es la señal medida (grados), L es el paso óptico de la cubeta (cm), [c] es la concentración de proteínas y # corresponde con el número de enlaces peptídicos de la proteína (número de residuos - 1).

Análisis del contenido α-hélice en diferentes condiciones de solventes.

Titulaciones con TFE. Las titulaciones fueron realizadas barriendo un rango entre 0 y 50% TFE v/v e incubando 5 minutos entre punto y punto para que el sistema llegue al equilibrio. Las proteínas fueron medidas a una concentración de 20 µM y los péptidos de E7 fueron medidos dentro de un rango de concentración entre 20 y 80 µM según el tamaño del péptido.

Ajuste a equilibrio de dos estados coil-α-hélice. La elipticidad molar media a 222 nm, la cual se asume proporcional al contenido α-hélice, fue expresada en función del cociente [TFE]/[H_2O] y los datos fueron ajustados a un modelo de equilibrio de dos estados *coil*/hélice [108]. Este modelo asume que la energía libre de formación de α-hélice depende linealmente con el cociente [TFE]/[H_2O]. Para obtener los parámetros termodinámicos $\Delta G^{0,H2O}$ y el factor *m*, los valores de señal fueron ajustados a la siguiente ecuación:

[Ecuación 2]

$$[\theta] = \frac{[\theta]^{TFE} + [\theta]^{H2O} * \exp((-\Delta G^{0,H_2O} - m[H_2O]/[TFE])/RT)}{1 + \exp((-\Delta G^{0,H_2O} - m[H_2O]/[TFE])/RT)}$$

Donde $[\theta]^{H2O}$ y $[\theta]^{TFE}$, corresponden a los valores de elipticidad molar media en agua y en TFE alto, respectivamente, R es la constante de los gases y T es la temperatura en ºK.

Curvas de pH en 30% TFE. Para realizar las curvas de pH de los distintos fragmentos de E7N, se utilizó el buffer de amplió rango citrato-fosfato. Los distintos péptidos de E7N fueron incubados en 10 mM citrato-fosfato entre pH 3.0 y 8.0 con 30% TFE. Cada tubo fue incubado por una hora a temperatura ambiente, y previo a la medida de CD el valor de pH fue chequeado en un pHimetro con un micro-électrodo (Denver Instrument). *Análisis de datos*. Los valores de elipticidad molar media a 222 nm fueron ajustados a una variante de la ecuación de Henderson-Hasselbalch [147], a partir de la cual se obtuvieron los valores de los *pKa* globales para cada fragmento, asumiendo que los grupos titulables están desprotonados en el estado inicial y protonados en el estado final:

[Ecuación 3]

$$y = (D_{in} + D_m * pH) * f_D + (P_{in} + P_m * pH) * f_P$$

Donde *y* se corresponde con la señal medida, D_{in} y P_{in} son las señales de los estados desprotonados y protonados, respectivamente. D_m y P_m corresponden con las variaciones lineales de la señal de cada estado con el pH. Las fracciones de las especies protonadas y desprotonadas (f_P y f_D), fueron definidas por las siguientes ecuaciones:

[Ecuación 4]

$$f_D = \frac{10^{(pKa - pH)}}{1 + 10^{(pKa - pH)}}$$

[Ecuación 5]

$$f_P = \frac{1}{1 + 10^{(pKa - pH)}}$$

Estabilización de estructura secundaria por SDS. Los experimentos fueron realizados en buffer formiato de sodio 10 mM (pH 3.0). Los distintos péptidos del dominio E7N en concentraciones de entre 20 y 40 µM fueron incubados por una hora a temperatura ambiente en distintas concentraciones de SDS en condiciones sub y supramicelares (1mM y 25 mM). La concentración micelar crítica para estas condiciones experimentales fue de 5 mM [64].

Estimación del contenido α-hélice a partir de datos del ajuste o de señal de CD. La estimación del número de residuos en hélice a partir de los datos de CD así como también de los datos obtenidos a partir del ajuste, fue obtenida a partir de la siguiente ecuación [148]:

[Ecuación 6]

$$[\theta]_{222nm} = -39{,}500 * (1 - \frac{2.57}{n})$$

Esta relación empírica permite definir MRW esperado para una proteína 100% en conformación α-hélice. Donde n es el número de residuos de la proteína.

Determinación de estructura PII estabilizada por temperatura.

Los elementos de estructura PII presente en los distintos péptidos o fragmentos fueron estabilizados a bajas temperaturas. Para determinar la inducción de este tipo de estructura, fueron analizados los espectros diferencias entre (5-75 ºC) medidos por CD para cada una de las especies.

Experimentos de desnaturalización térmica.

Modelo reversible. La desnaturalización térmica reversible de P y sus fragmentos fue seguida por el cambio en la elipticidad a 222 nm entre 5 y 85 ºC, en 20 mM Tris-HCl (pH 7.5), 0.2 M NaCl o en 20 mM Formiato de Sodio (pH 3.0), 0.2 M NaCl, según sea indicado. Las medidas fueron realizadas con una velocidad de barrido de 5 ºC/min.

Los cambios en elipticidad a 222 nm en función de la temperatura pueden ser descriptos por la siguiente ecuación:

[Ecuación 7]

$$Signal = \frac{(U_0 + U_m \cdot (T - 273.15) \cdot e^{-\Delta G/RT})}{1 + e^{-\Delta G/RT}} + \frac{(N_0 + N_m \cdot (T - 273.15))}{1 + e^{-\Delta G/RT}}$$

Donde U_0 y N_0, corresponden a la elipticidad molar media de los estados desplegados y nativos, respectivamente. U_m y N_m son parámetros que tienen en cuenta la variación lineal de la señal con la temperatura. ΔG el la energía libre de desplegado, R es la constante de los gases expresada como 1.9872 $calK^{-1}mol^{-1}$ y T es la temperatura en ºK. Como no hay cambios en la molecularidad durante la transición térmica, los datos fueron ajustados a la ecuación de Gibbs-Helmholtz que describe los cambios en la transición de desplegado de una proteína monomérica en función de la temperatura [149]:

[Ecuación 8]

$$\Delta G = \Delta H(1 - T/T_m) - \Delta C_p(T_m - T) + T\ln(T/T_m)$$

Donde ΔH es la entalpía de desplegado, ΔCp es la capacidad calorífica de desplegado y T_m es el punto medio de la transición de desplegado en ºK. Los datos de los experimentos de desplegado térmico fueron ajustados a las ecuaciones 7 y 8, considerando ΔCp =0.

Nota: para los experimentos de desnaturalización térmica se probaron también buffer 20 mM fosfato de sodio (pH 7.5), 50 mM NaCl; así como también distintas concentraciones de sal 50 nM NaCl a 200 nM NaCl y de proteína 80 µM a 10 µM, sin observar ninguna diferencia en los valores de T_m determinados. También se evaluó utilizar una velocidad de barrido de 1 ºC/min, sin observar diferencias.

Experimentos de desnaturalización química en Cloruro de Guanidinio (Gdm.Cl).

Las curvas de desplegado químico barriendo un rango entre 0 y 7,5 M Gdm.Cl fueron medidas en buffer 50 mM Tris-Cl (pH 7.5), 20 mM NaCl en tubos separados incubados a temperatura ambiente por 16 horas, según [132]. Para los análisis de la primera transición de desplegado (0 y 4 M Gdm.Cl) se realizaron titulaciones en buffer 50 mM Tris-Cl (pH 7.5) y 20 mM NaCl, fueron medidas a 20 y 10 ºC, con una incubación entre punto y punto de 5 minutos (tiempo suficiente para que el sistema llegue al equilibrio).

Experimentos cinéticos de desplegado.

Los experimentos cinéticos de desplegado térmico y químico fueron realizados siguiendo el cambio de la señal de la banda de CD a 222 nm en el tiempo. Las medidas fueron realizadas a una concentración de 10 µM de P o 20 µM P_C, en buffer 20 mM Tris-HCl (pH 7.5), 50 mM NaCl.

Cinética de desplegado térmico. Fue realizada midiendo saltos de temperatura entre 10 y 40 ºC, o viceversa. La proteína P o el fragmento P_C provenientes de un *stock* concentrado equilibrado a 10 o 40 ºC, fueron transferidos a un buffer equilibrado a 40 o 10 ºC en el equipo de CD, respectivamente. Nota: se utilizaron distintas cubetas de CD (0.2 a 1 cm de paso óptico) para evaluar posibles efectos de la distintos volúmenes de mezcla.

Cinética de desplegado químico. La proteína P o el fragmento P_{NTET} ambos provenientes de un *stock* concentrado equilibrado a 10 ºC fueron transferidos a una cubeta de CD en el equipo conteniendo *buffe*r y distintas concentraciones de Gdm.Cl (0.5-6M), todo a 10 ºC. La transición lenta observada para la cinética a 6M Gdm.Cl fue ajustada a una función exponencial simple.

[Ecuación 9]

$$y = A * e^{-kt}$$

Donde *A* es la amplitud de la transición, *t* es el tiempo en minutos y *k* es la constante de velocidad de desplegado.

Nota: para ambos experimentos de desplegado, las líneas de base fueron monitoreadas por 2 minutos antes de realizar el mezclado entre el *buffer* y la proteína.

VII-ESTUDIOS CONFORMACIONALES POR ESPECTROSCOPÍA DE FLUORESCENCIA.

Análisis de la unión a ANS.

Titulaciones. Espectros de emisión de fluorescencia entre 400 y 600 nm fueron medidos en un espectrofluorimetro JASCO FP-6500 con una longitud de onda de excitación de 370 nm, una velocidad de barrido de 100 nm/min y un ancho de banda de excitación y emisión de 5 nm, a 20 ºC. Titulaciones al equilibrio con ANS fueron realizadas para P, P_{TETC} y P_{NTET}, con tiempos de incubación entre punto y punto de 5 minutos. *Medidas a concentración fija de ANS.* Los distintos fragmentos de P fueron incubados por 20 minutos con 100 µM ANS.

Desplazamiento de ANS por Gdm.Cl.

El desplazamiento del ANS unido por parte del Gdm.Cl, fue evaluado por titulaciones del complejo proteína:ANS con concentraciones crecientes de Gdm.Cl incubando 5 minutos entre punto y punto.

VIII-DETERMINACION DE LAS PROPIEDADES HIDRODINAMICAS DE PEPTIDOS Y PROTEINAS.

Experimentos de difusión por PFG-NMR.

Las medidas de RMN presentadas en esta tesis fueron realizadas en colaboración con la Dra. Mariana Gallo, en un equipo de resonancia magnética nuclear Avance III Bruker 600 MHz, equipado con una criosonda de 5 mm de triple resonancia. Las medidas de difusión se realizaron mediante gradiente de campo pulsado (PFG, por sus siglas en inglés), utilizando una secuencia de pulsos PFG-SLED. La muestra de E7N fue preparada en una concentración de 1 mg/ml en D_2O a pH 7.5 o pH 5.0, con 10 mM TCEP. Se utilizó como estándar interno una solución de dioxano 0.03% v/v [34]. Todas las muestras de E7N y el estándar fueron preparadas en un volumen final de 300 µl y 2% de H_2O. El pH de la muestra fue controlado previa y posteriormente a la determinación en un pHimetro con un micro-electrodo (Denver Instrument).

La duración de los pulsos y tiempos de retardo en la secuencia se mantuvieron constantes y una serie de espectros fueron adquiridos variando la fuerza del gradiente de difusión entre 5% y 95%. El tiempo de pulso entre los gradientes fue de 4 mseg y la duración del retardo de difusión fue calibrada para cada muestra con el fin de obtener un decaimiento máximo de 80-90 % para las señales de la proteína y del dioxano entre el experimento con menor y mayor fuerza del gradiente (150 mseg y 20 mseg para E7N y dioxano, respectivamente). La intensidad de decaimiento para E7N fue homogénea en todo el espectro del protón. En consecuencia, se ajustaron las intensidades del decaimiento de la señal de E7N a través de las regiones alifáticas y aromáticas. En los experimentos de dioxano se aplicó un filtro T_2 para medir selectivamente la señal del dioxano sin interferencia de la señal de la proteína, a fin de disminuir el error experimental especialmente en los experimentos con valores altos de gradiente.

Finalmente, los valores de R_S para E7N en las diferentes condiciones experimentales fueron calculados a partir de la siguiente ecuación [34]:

[Ecuación 14]

$$R_S = (D_{diox} / D_{E7(1-40)}) * R_{Sdiox}$$

Donde, D_{diox} y D_{E7N} corresponden a las medidas de los coeficientes de difusión del dioxano y de E7N, respectivamente y R_{Sdiox} al radio hidrodinámico efectivo del dioxano (2.12 Å) [34].

Ultracentrifugación analítica (AUC).

Los experimentos de ultracentrifugación analítica (AUC, por sus siglas en inglés) fueron realizados por el método de velocidad de sedimentación (SV) en colaboración con el Dr. Andrés Salvay de la Universidad Nacional de La Plata. Se utilizó una ultracentrífuga analítica Beckman Coulter XL-I, a una velocidad de 50.000 rpm con un rotor de 8 huecos ANTi-50. Todas las celdas fueron equipadas con ventanas de safiro y con centros de paso óptico de 0.3 cm. Se utilizó un volumen de 100 µl de muestra, así como también de solvente de referencia. Los perfiles de velocidad de sedimentación fueron adquiridos a intervalos de 13 minutos entre celdas durante 24 horas utilizando ópticas de absorbancia. La densidad y viscosidad del buffer, así como también el volumen parcial específico de cada especie fue calculado utilizando el programa SEDNTERP. Los análisis de los experimentos de SV fueron realizados mediante el modelo de distribución continua y de especies no interactuantes *c(s)* del programa SEDFIT. Se evaluaron muestras de E7N, E7(1-20) y E7(16-40) a concentraciones de 1 mg/ml, 0.5 mg/ml y 0.33 mg/ml.

Los coeficientes de sedimentación s_0 y difusión D_{0app}, fueron derivados de la aproximación lineal a dilución infinita. Dichos coeficientes fueron utilizados para obtener los valores correspondientes de la masa molecular aparente (M_{app}) y el radio hidrodinámico (R_S) para cada péptido, a partir de la ecuación de Stokes-Einstein y Svedberg:

[Ecuación 10]

$$R_S = RT/(N_A 6\pi\eta^{\circ} D_0)$$

[Ecuación 11]

$$M_{app} = RT * s_0/(D_0(1-\rho^{\circ} v))$$

Donde *R* corresponde a la constante de los gases, T a la temperatura absoluta, N_A al número de Avogadro y η a la viscosidad del solvente [33]. Se determinó el parámetro *f/fmin,* como el cociente entre el R_S y el R_{min} que se describe como el R_S teórico considerando un estado de no hidratación.

[Ecuación 12]

$$f/f_{\min} = R_S / R_{\min}$$

R_{min} depende de la raíz cubica de la masa molar (M) y del volumen no hidratado de la partícula (V).

[Ecuación 13]

$$\mathrm{V} = (4/3)\pi \mathrm{R}^3_{\min} = \mathrm{M} * v/\mathrm{N_A}$$

Análisis hidrodinámico por cromatografía de exclusión molecular (SEC).

Los experimentos de SEC fueron realizados en una columna Superdex S200 HR 10/300 (24 ml) (GE Helthcare), equilibradas en 20 mM Tris-HCl (pH 7.5) y 0.2 M NaCl o 20 mM Formiato

de Sodio (pH 3.0) y 0.2 M NaCl, a temperatura ambiente. Las muestras fueron incubadas por 1 hora y centrifugadas a 13.000 rpm antes de ser inyectadas en la columna. La misma fue realizada en un equipo de FPLC siguiendo la señal de absorbancia a 220 nm.

Para la calibración de la columna se utilizaron estándares de peso molecular que corresponden a las proteínas globulares: Tiroglobulina (660 kDa), Ferritina (440 kDa), Catalasa (232 kDa), BSA (67kDa), MBP (45kDa), Ribonucleasa A (13.4kDa). El volumen de muerto (V_0) y el volumen total (V_T) fueron determinados inyectando el polímero Azul Dextrano y acetona, respectivamente para cada condición de buffer. A partir de los volúmenes de elución (V_e) de cada especie, se calculó el coeficiente de partición de las muestras en la columna como $K_{av}=(V_e-V_0)/(V_T-V_0)$.

Medidas de dispersión de luz estática (SLS).

Los pesos moleculares de P y de los distintos sub-fragmentos fueron determinados por dispersión de luz estática (SLS, por sus siglas en inglés) en un Instrumento Precision Detectors PD2010, conectado en línea a un equipo de FPLC y a un sistema de refractómetro diferencial LKB 2142. Se inyectaron 200 µl de las distintas especies de P a una concentración de 1 mg/ml, en una columna Superdex S200 HR 10/300 (24 ml) (GE Helthcare) equilibradas en 20 mM Tris-HCl (pH 7.5) y 0.2 M NaCl. La señal de dispersión de luz a 90° y el índice de refracción fueron analizados utilizando el software Discovery32 (Precision Detectors). Para calibrar la medición se utilizó BSA (PM: 66.5 kDa). Las muestras fueron incubadas durante 1 hora a temperatura ambiente en el buffer de corrida y centrifugadas antes de ser inyectadas.

IX-ESTUDIOS ESTRUCTURALES DE RESOLUCIÓN ATÓMICA POR RMN.

Breve descripción teórica. En esta técnica, pulsos de radiofrecuencia en el rango de los MHz, son aplicados a una muestra sometida a un campo magnético homogéneo (B_0). En presencia de dicho campo magnético externo, los estados de *spin* de los núcleos presentan variaciones energéticas y sus frecuencias (ω) son descriptas por la ecuación de Bloch ($\omega = -\gamma B_0$) (Block). Donde γ está determinada por la receptividad del tipo de núcleo para la espectroscopía, y por un parámetro B_0 que es comúnmente denominado potencia del campo magnético externo. Sin embargo, núcleos químicamente iguales presentan diferencias en las frecuencias observadas debido a que se encuentran en entornos distintos. Estas desviaciones respecto de la predicción de Bloch, se denominan desplazamiento químico (CS, por sus siglas en inglés) y pueden expresarse mediante la siguiente ecuación: $\omega = \gamma (1-\delta) B_0$, donde ($\delta$) corresponde al desplazamiento químico que modifica a γB_0 en una magnitud muy pequeña pero observable (partes por millón, ppm). El fenómeno de CS se debe a que los movimientos de los electrones

inducidos por el campo magnético externo genera campos magnéticos secundarios. En el entorno de las proteínas, las diferentes estructuras secundarias representan distintos ambientes y por lo tanto un campo magnético resultante, particular para ese núcleo. Existen desviaciones sistemáticas y particulares para cada estructura canónica (α-hélice o lámina-β) y por lo tanto, se puede comparar la desviación que presenta un núcleo dado, respecto del valor esperado para ese mismo núcleo, por ejemplo en el estado de *random coil*.

Asimismo, los efectos de interacción entre las señales de resonancia de dos núcleos, pueden estar mediados por los electrones que forman el enlace químico entre los mismos y recibe el nombre de "acoplamiento escalar". Este fenómeno depende del ángulo dihedro que formen los núcleos interactuantes. La fuerza de la interacción se puede medir utilizando al parámetro $^{n}J_{ab}$, "constante de acoplamiento escalar", donde n designa el número de uniones covalentes que separa los dos núcleos, a y b. La magnitud de $^{n}J_{ab}$ se expresa en hertz (Hz) y n puede tomar valores entre 1 y 4.

Por otra parte, a diferencia del acoplamiento escalar, existe otro tipo de interacción que no posee carácter mixto nuclear-electrónico. La interacción magnética dipolar sólo depende del intercambio energético entre dos *spin* que se encuentran cercanos en el espacio. Ambos núcleos no se relajan independientemente si éstos se encuentran a distancias menores a 5 Å y este fenómeno recibe el nombre de NOE (Nuclear Overhauser effect). De esta manera, las restricciones espaciales para la determinación estructural se derivan principalmente de los NOE. Para el estudio de los NOE se realizan experimentos de relajación cruzada y la secuencia de pulsos que se utiliza recibe el nombre de NOESY. Así, pueden asignarse secuencialmente polipéptidos sin marcar utilizando los NOE para correlacionar a los protones que se encuentran cercanos en el espacio.

Los experimentos de correlación heteronuclear ^{13}C-^{1}H fueron realizados en abundancia natural. Los gradientes de campo pulsado fueron aplicados para suprimir la señal del solvente y posibles artefactos espectrales. La detección de cuadratura en las dimensiones detectadas indirectamente fueron obtenidas usando States-TPPI (en las bidimensionales ^{1}H-^{1}H) o el método echo-antiecho (en las bidimensionales ^{1}H-^{13}C). Los datos fueron procesados con el software NMRPipe y analizados utilizando el software NMRViewJ.

Asignación de los desplazamientos químicos.

Para poder explorar las propiedades conformacionales de E7N, fueron asignadas todas las resonancias del ^{13}C y del ^{1}H del péptido E7(1-40) en abundancia natural utilizando una concentración de 3 mM, a 20 ºC. Las asignaciones fueron realizadas en los siguientes solventes: (i) H_2O pH 5.0 y pH 7.5 conteniendo 5% D_2O, (ii) 50% TFE-d_2 (Sigma-Aldrich) en H_2O pH 7.5 y (iii) 12% TFE-d_2 (Sigma-Aldrich) en H_2O a pH 5.0. Las muestras fueron tratadas con 10 mM TCEP con el fin de evitar la oxidación de la Cys 24 de E7N. El estado de oxidación

del péptido fue determinado antes y después de cada experimento por MALDI-TOF. No se detectaron precipitados o variación del pH de las muestras durante las mediciones.

Las asignaciones de los desplazamientos químicos del ^{13}C y del ^{1}H fueron realizadas utilizando una combinación de experimentos bidimensionales homo-nucleares y hetero-nucleares. ^{1}H-^{1}H NOESY (100 y 250 mseg de tiempos de mezclado); ^{1}H-^{1}H TOCSY (30 y 70 mseg de tiempo de mezclado); ^{1}H-^{13}C HMQC optimizado para observar las regiones alifáticas o aromáticas del espectro y HMQC-TOCSY (70 mseg de tiempo de mezcla). Los experimentos NOESY fueron adquiridos con 4096 (t2) y 1024 (t1) puntos complejos, con ventanas espectrales de 9615 y 6000 Hz en la dimensión directa e indirecta, respectivamente. Los *sets* de datos de TOCSY consistieron en 2048 (t2) y 800 (t1) puntos complejos con la misma ventana espectral utilizada en los experimentos de NOESY. Los datos para los experimentos ^{1}H-^{13}C HMQC y los ^{1}H-^{13}C HMQC-TOCSY fueron colectados con 2048 (t2) y 800 (t1) puntos complejos, y 9615 (^{1}H) y 8906 (^{13}C) Hz de ventana espectral, respectivamente. El *carrier* para el ^{13}C fue centrado en 40.00 ppM. Para los ^{1}H-^{13}C HMQC aromáticos, los puntos complejos 2048 (t2) y 256 (t1) fueron colectados, con ventanas espectrales de 9615 y 3774 Hz, respectivamente, con el *carrie*r de carbono centrado a 130.00 ppm.

Constantes de acoplamiento escalar. Las constantes de acoplamiento escalar $^{1}JCH_{\alpha}$ fueron medidas a través de experimentos ^{1}H-^{13}C-HMQC sin desacoplamiento en la dimensión del protón. Estos experimentos fueron realizados en agua pH 7.5 (5% D_2O) a 4 ºC, agua pH 5.0 a 20 ºC y agua con 50% TFE-d2 (Sigma-Aldrich) a 20 ºC. Para aumentar la resolución de la medida, la región alfa fue seleccionada y el resto de las señales fueron plegadas en la ventana espectral reducida. Los ^{1}H-^{13}C HMQCs fueron adquiridos con el *carrier* de ^{13}C centrado a 40.00 ppm, pero con puntos complejos 2048 (t2) y 2048 (t1) y ventanas espectrales de 9615 (^{1}H) y 2868 (^{13}C) Hz, respectivamente.

X-HERRAMIENTAS COMPUTACIONALES Y PROGRAMAS UTILIZADOS.

El procesamiento y el ajuste de todos los datos fue realizado utilizando el programa PROFIT (Quantumsoft, Xurich, Switzerland). Los alineamientos de secuencias fueron realizados utilizando el programa Seaview. Las visualizaciones y el análisis de estructuras bajadas del Protein data bank (pdb) fueron realizadas utilizando Pymol (https://www.pymol.org). El análisis de tendencia a adoptar estructura α-hélice fue realizado utilizando el programa AGADIR (http://agadir.crg.es). Las secuencias de las proteínas fueron obtenidas de Uniprot. Los análisis estadísticos de ANOVA de una vía y el test de STUDENT fueron realizados utilizando el programa GraphPad Prism.

Los logos de secuencia fueron construidos con el programa WebLogo y Logobar (Schneider and Stephens, 1990) (http://weblogo.berkeley.edu/logo.cgi) y utilizando un alineamiento manual de las secuencias (Figura apéndice A13). Elegimos representar el alineamiento múltiple de secuencias de P de la subfamilia *Pneumovirinae* con el visualizador LogoBar [140]. El LogoBar como representación gráfica es muy similar al logo de secuencia obtenido por el programa WebLogo [122], pero a diferencia de letras utiliza apilamientos de barras para cada posición de la secuencia, permitiendo analizar perfiles de gran longitud. La altura total de cada apilamiento indica la conservación de secuencia en esa posición (medida en bits), mientras que la altura de cada barra dentro de un apilamiento refleja la frecuencia relativa del aminoácido correspondiente [122, 140]. Adicionalmente, el LogoBar grafica las regiones de brechas del alineamiento o huecos (*gap,* en inglés), permitiendo apreciar junto con la variabilidad de secuencia producto de mutaciones de tipo sustituciones, las mutaciones fruto de inserciones o deleciones entre las distintas secuencias aminoacídicas alineadas [140].

Cálculo de gráficos carga vs hidrofobicidad (CH plots). Los gráficos de carga versus hidrofobicidad (CH) fueron generados según [26]. La carga neta a pH 7.0 fue calculada utilizando la herramienta *on line* de innovagen "Peptide Property Calculator" (http://pepcalc.com/ppc.php). Esta herramienta determina la carga neta (Z) a partir de la siguiente expresión que describe la carga neta a un dado pH como:

[Ecuación 15]

$$Z = \sum_i N_i \frac{10^{\mathrm{pKa}_i}}{10^{\mathrm{pH}} + 10^{\mathrm{pKa}_i}} - \sum_j N_j \frac{10^{\mathrm{pH}}}{10^{\mathrm{pH}} + 10^{\mathrm{pKa}_j}}$$

donde N_i corresponde al número, los valores pKa_i y pKa_i, de los residuos Arg, Lys, His y el extremo N terminal de la proteína. Y N_j corresponde al número, los valores pKa_j y pKa_j, de los residuos Asp, Glu, Cys, Tyr y el extremo C terminal de la proteína. Luego, la carga neta media fue obtenida como el modulo de Z dividido el número de residuos de la proteína.

Las hidrofobicidades individuales fueron calculadas utilizando el programa Protscale de Expasy (http://www.expasy.ch/tools), usando la opción Hphob/Kyte & Doolittle'', con una ventana de 5 residuos y pidiendo que muestre los valores normalizados entre 0 y1. Esta herramienta da los valores normalizados de los índices de hidrofobilcidad para cada residuo descripto por *Kyte y Doolittle* [150]. Luego, se suma la hidrofobicidad normalizada para cada residuo y la hidrofobicidad media se obtiene dividiendo dicho valor de la hidrofobicidad media por el número de residuos de la proteína.

APENDICE

Figura A1. Mapa y detalle de secuencia del vector pRSETA modificado *ad hoc.* A-Mapa de secuencia del vector pRSETA, en color se marcan los sitios BamHI y EcoRI del sitio múltiple de clonado, así como también la secuenca LVPRGS correspondiente al sitio de corte por trombina. B-Detalle de secuencia del vector. En rosa se indica el sitio de corte para las enzimas NdeI y BamHI que flanquean al *tag* de histidina.

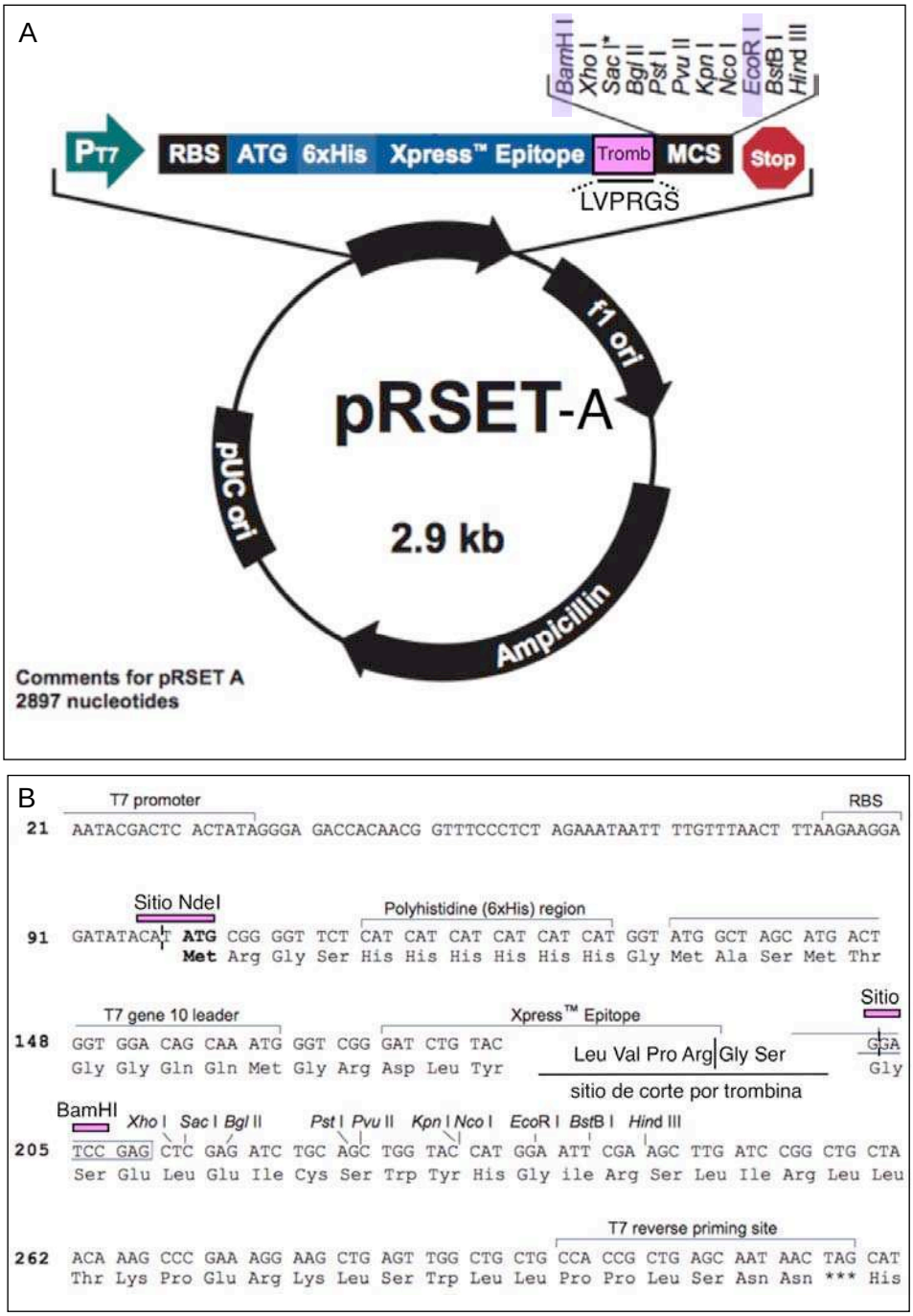


Figura A2. Alineamiento de secuencia de P de RSV con las variantes pTrombina I y II, indicando los sitios donde se realizó la inserción del sitio de corte.

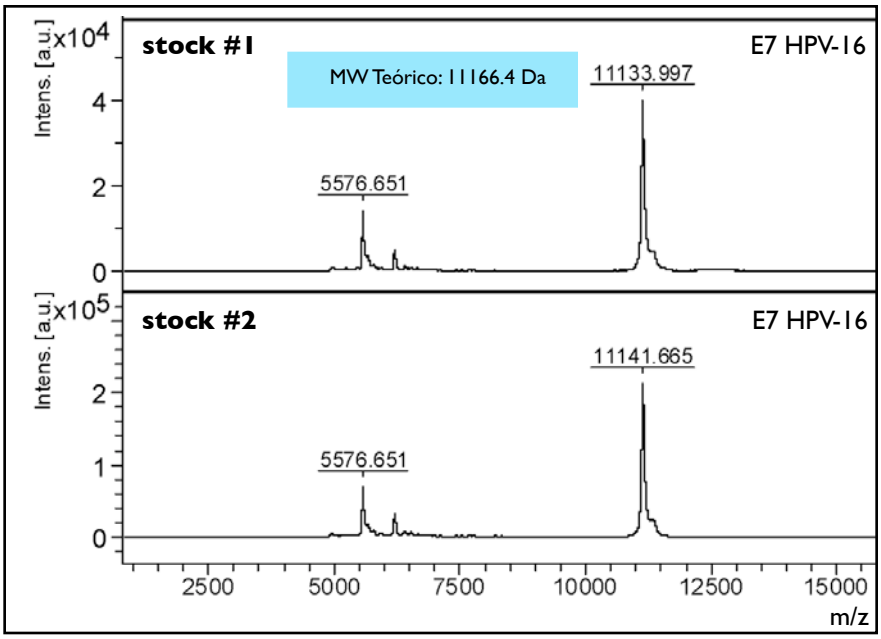
Vector pRSET-A(PTrombina-I)

```
CLUSTAL 2.1 multiple sequence alignment


Ptrombina-I     MEKFAPEFHGEDANNRATKFLESIKGKFTSPKDPKKKDSIISVNSIDIEVTKESPITSNS 60
P               MEKFAPEFHGEDANNRATKFLESIKGKFTSPKDPKKKDSIISVNSIDIEVTKESPITSNS 60
                ************************************************************
                                                   Inserción
Ptrombina-I     TIINPTNETDDTAGNKPNYQRKPLVSFKEDPTPSDNPFSKLYLVPRGSKETIETFDNNEE 120
P               TIINPTNETDDTAGNKPNYQRKPLVSFKEDPTPSDNPFSKLY------KETIETFDNNEE 114
                ****************************************      ************

Ptrombina-I     ESSYSYEEINDQTNDNITARLDRIDEKLSEILGMLHTLVVASAGPTSARDGIRDAMIGLR 180
P               ESSYSYEEINDQTNDNITARLDRIDEKLSEILGMLHTLVVASAGPTSARDGIRDAMIGLR 174
                ************************************************************

Ptrombina-I     EEMIEKIRTEALMTNDRLEAMARLRNEESEKMAKDTSDEVSLNPTSEKLNNLLEGNDSDN 240
P               EEMIEKIRTEALMTNDRLEAMARLRNEESEKMAKDTSDEVSLNPTSEKLNNLLEGNDSDN 234
                ************************************************************

Ptrombina-I     DLSLEDF 247
P               DLSLEDF 241
                *******
```

Vector pRSET-A(PTrombina-II)

```
CLUSTAL 2.1 multiple sequence alignment


ptrombinaII     MEKFAPEFHGEDANNRATKFLESIKGKFTSPKDPKKKDSIISVNSIDIEVTKESPITSNS 60
P               MEKFAPEFHGEDANNRATKFLESIKGKFTSPKDPKKKDSIISVNSIDIEVTKESPITSNS 60
                ************************************************************

ptrombinaII     TIINPTNETDDTAGNKPNYQRKPLVSFKEDPTPSDNPFSKLYKETIETFDNNEEESSYSY 120
P               TIINPTNETDDTAGNKPNYQRKPLVSFKEDPTPSDNPFSKLYKETIETFDNNEEESSYSY 120
                ************************************************************
                                                   Inserción
ptrombinaII     EEINDQTNDNITARLDRIDEKLSEILGMLHTLVVASAGPTSLVPRGSARDGIRDAMIGLR 180
P               EEINDQTNDNITARLDRIDEKLSEILGMLHTLVVASAGPT------SARDGIRDAMIGLR 174
                ****************************************      **************

ptrombinaII     EEMIEKIRTEALMTNDRLEAMARLRNEESEKMAKDTSDEVSLNPTSEKLNNLLEGNDSDN 240
P               EEMIEKIRTEALMTNDRLEAMARLRNEESEKMAKDTSDEVSLNPTSEKLNNLLEGNDSDN 234
                ************************************************************

ptrombinaII     DLSLEDF 247
P               DLSLEDF 241
                *******
```

Figura A3. Espectro de MALDI-TOF de E7 y E7(27-98) de HPV-16.



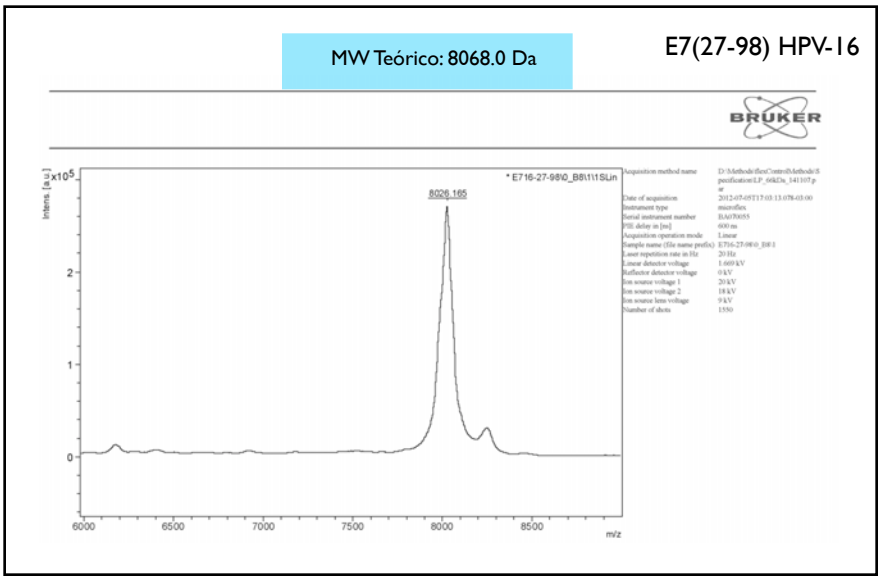


Figura A4. Espectro de MALDI-TOF, perfil de cromatográfico y SDS-PAGE 15% de P de RSV. Medidas mostrando la calidad de las preparaciones de P utilizada para los experimentos del capitulo 2. Se muestran espectros de MALDI-TOF de tres stocks independientes y un perfil de exclusión molecular en Superdex S200 analítica y un gel SDS-PAGE 15% representativo.

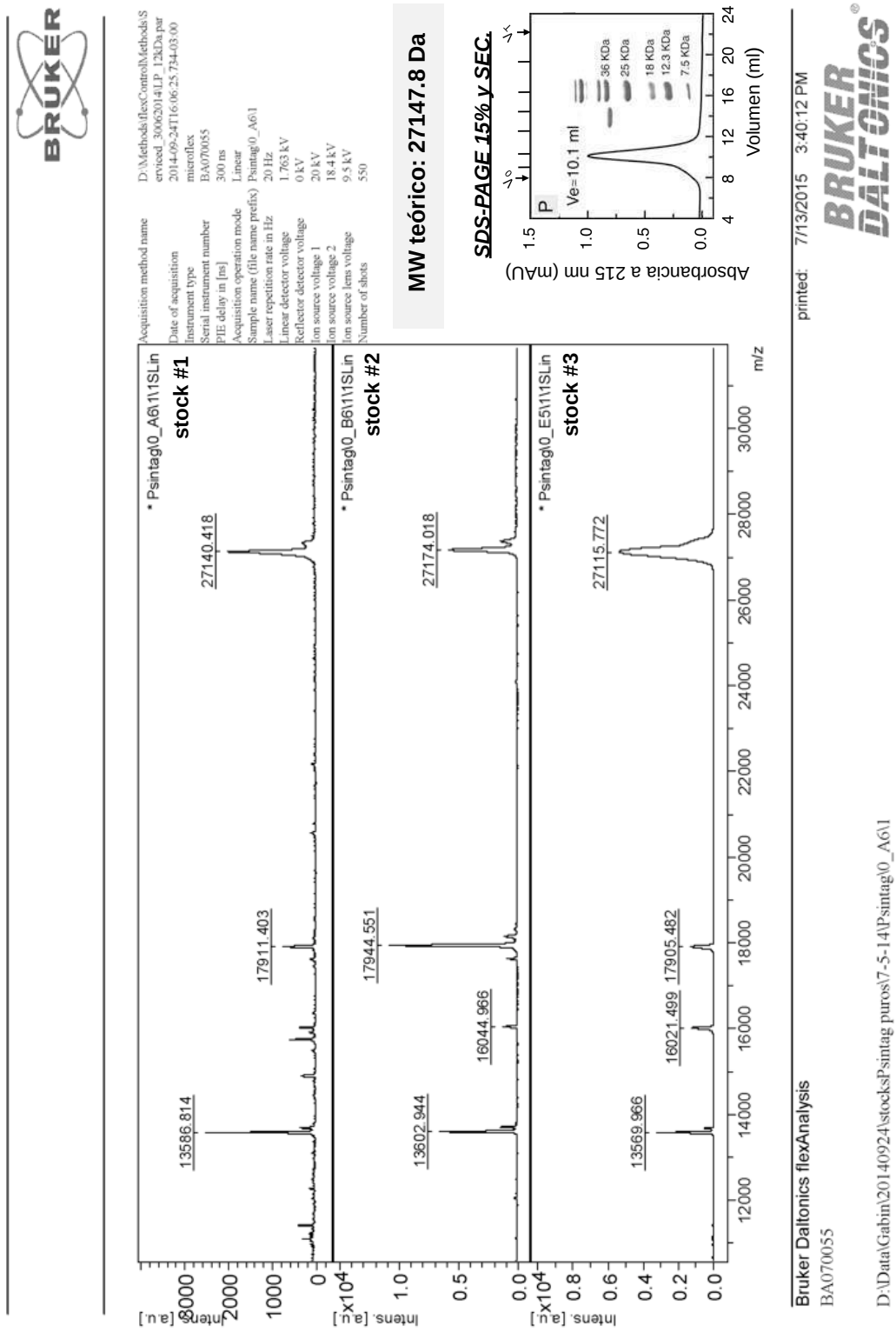


Figura A5. Espectro de MALDI-TOF, perfil de cromatográfico y SDS-PAGE 15% del fragmento P_{NTET}. Medidas mostrando la calidad de las preparaciones de P_{NTET} utilizada para los experimentos del capitulo 2. Se muestran espectros de MALDI-TOF de tres stocks independientes y un perfil de exclusión molecular en Superdex S200 analítica y un gel SDS-PAGE 15% representativo.

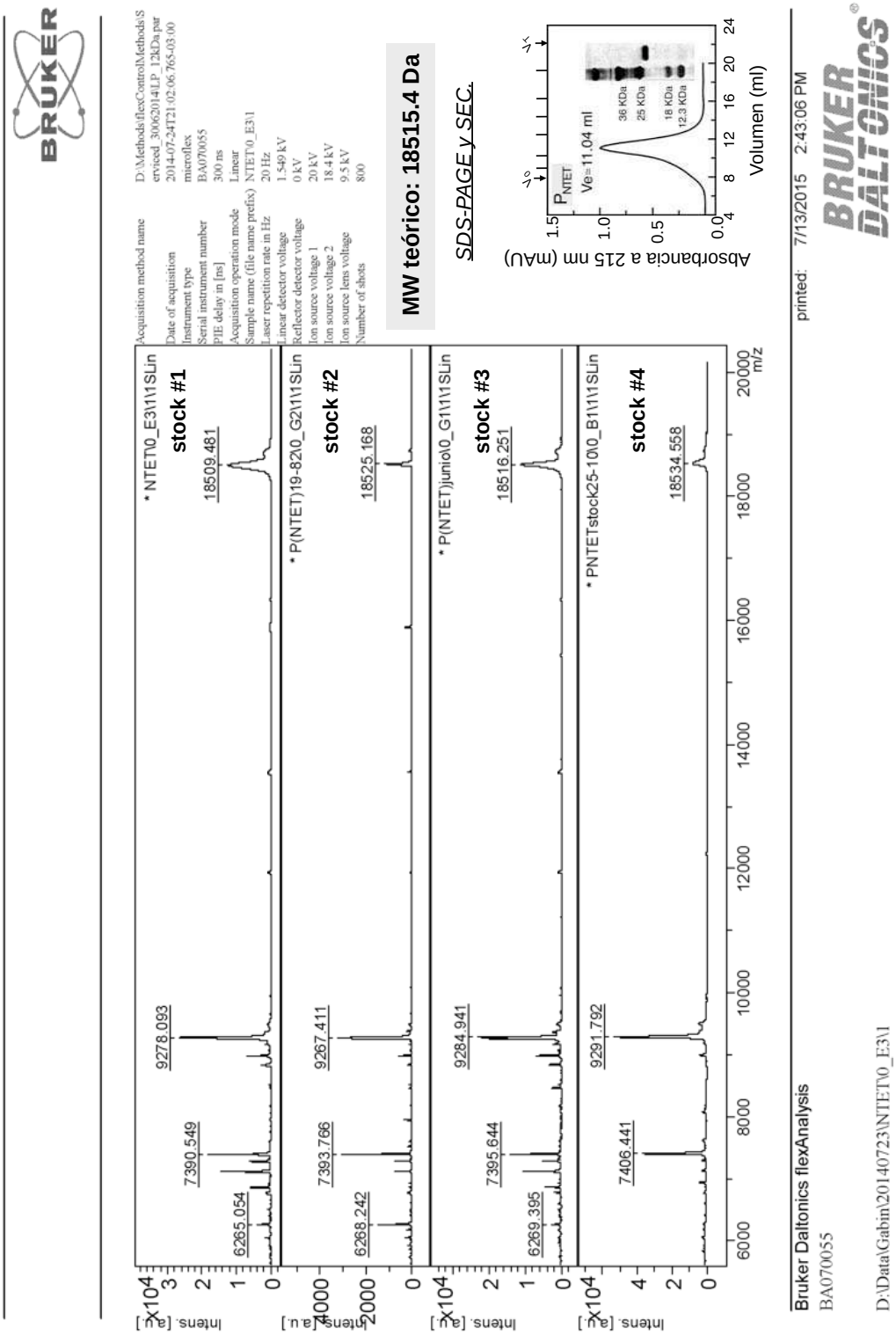


Figura A6. Espectro de MALDI-TOF, perfil de cromatográfico y SDS-PAGE 15% del fragmento P_{TETC}. Medidas mostrando la calidad de las preparaciones de P_{TETC} utilizada para los experimentos del capitulo 2. Se muestran espectros de MALDI-TOF de tres stocks independientes y un perfil de exclusión molecular en Superdex S200 analítica y un gel SDS-PAGE 15% representativo.

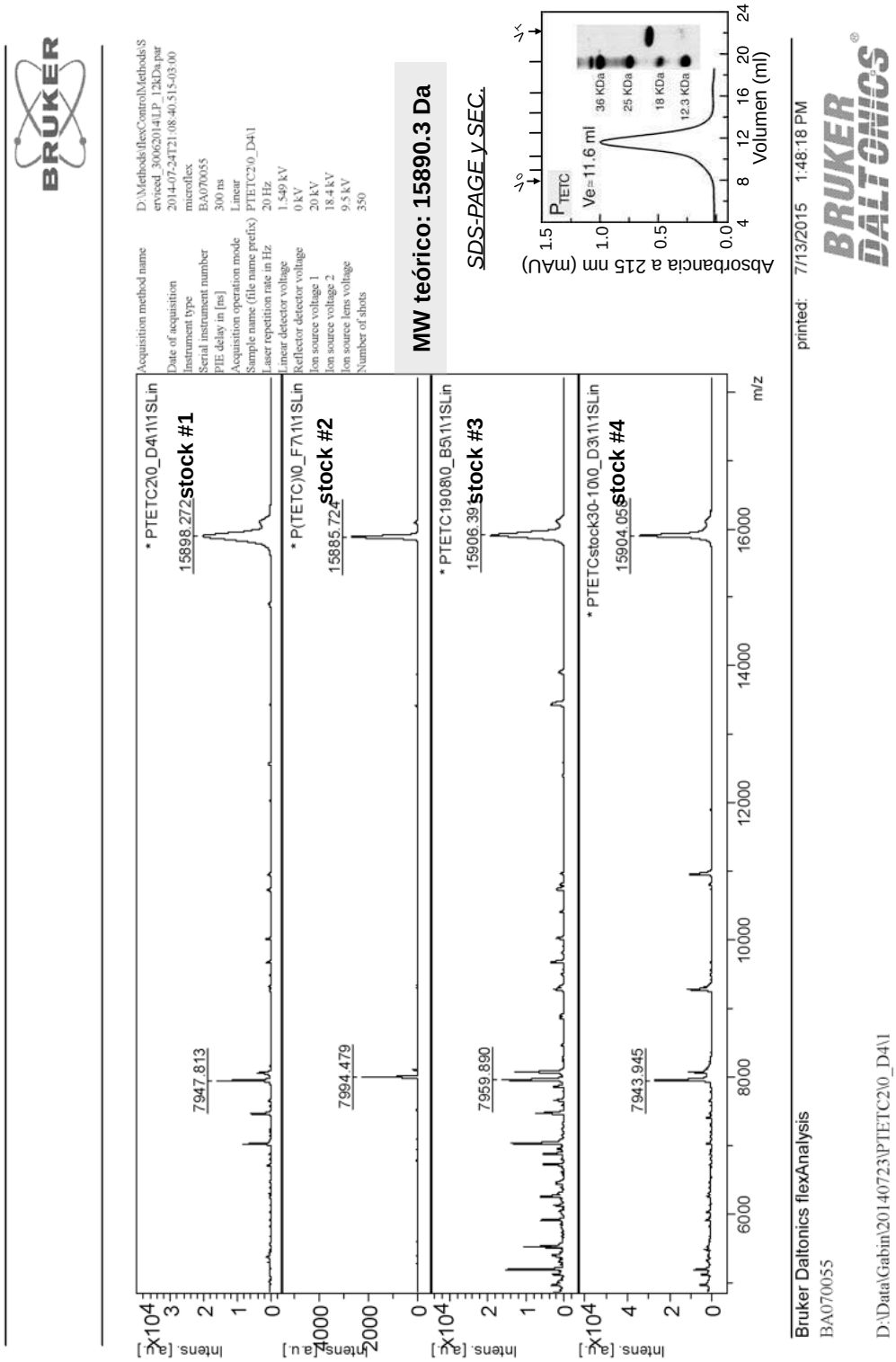


Figura A7. Espectro de MALDI-TOF, perfil de cromatográfico y SDS-PAGE 15% del fragmento P_N. Medidas mostrando la calidad de las preparaciones de P_N utilizada para los experimentos del capitulo 2. Se muestran espectros de MALDI-TOF de tres stocks independientes y un perfil de exclusión molecular en Superdex S200 analítica y un gel SDS-PAGE 15% representativo.

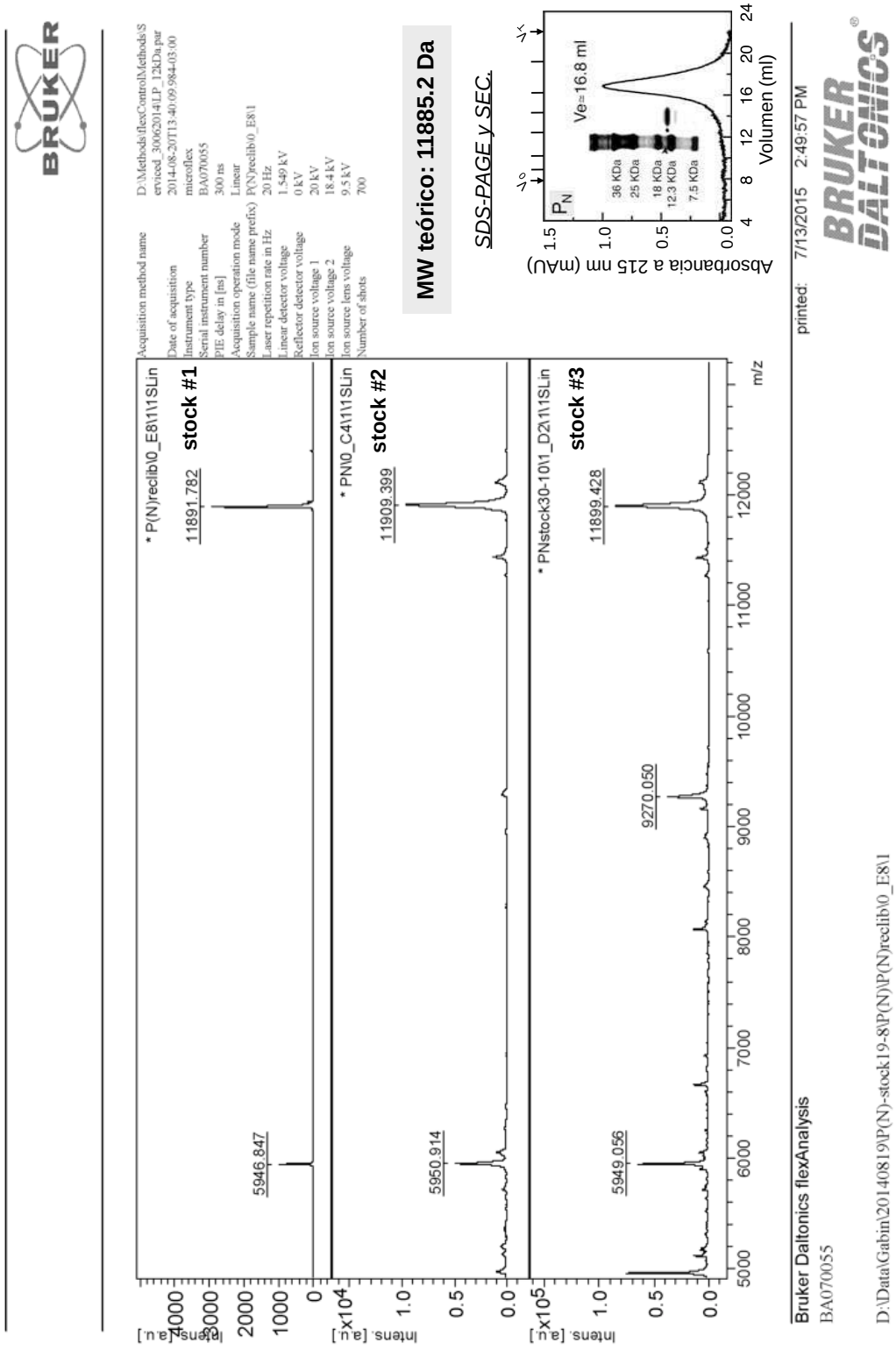


Figura A8. Espectro de MALDI-TOF, perfil de cromatográfico y SDS-PAGE 15% del fragmento P_C. Medidas mostrando la calidad de las preparaciones de P_C utilizada para los experimentos del capitulo 2. Se muestran espectros de MALDI-TOF de tres stocks independientes y un perfil de exclusión molecular en Superdex S200 analítica y un gel SDS-PAGE 15% representativo.

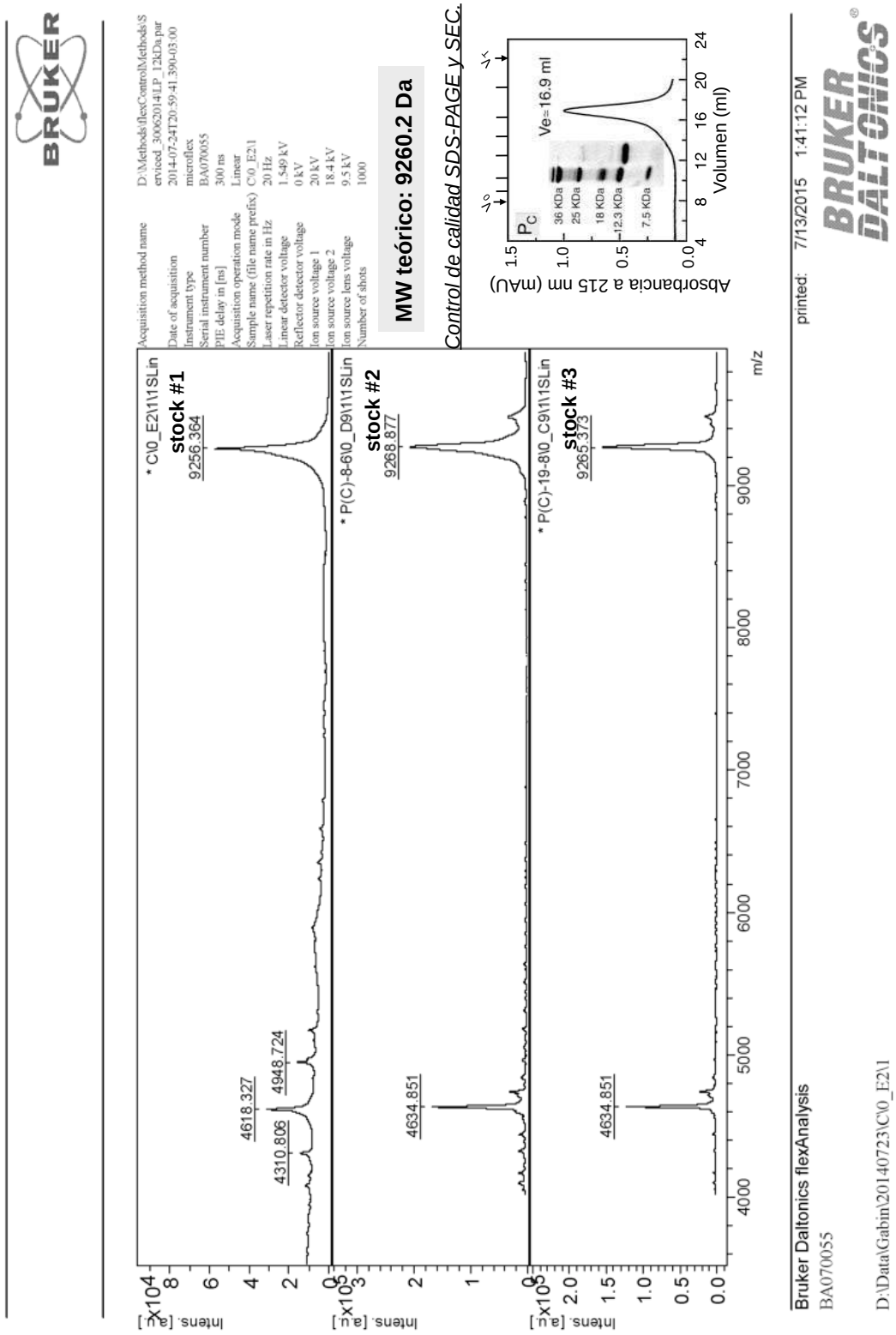


Figura A9. Obtención por proteólisis tríptica del subfragmento P_{C*}. Proteólisis del fragmento P_C seguida por HPLC y espectrometría de masas. Se muestra la secuencia de uno de los fragmentos de proteólisis predicho por el programa Mass Peptide de Expasy. El mismo concuerda al 100% con el pico identificado por HPLC y MALDI.

Segun Mass Peptide (Expasy)

4958.2398 Da LRNEESEKMAKDTSDEVSLN PTSEKLNNLLEGNDSDNDLS LEDF

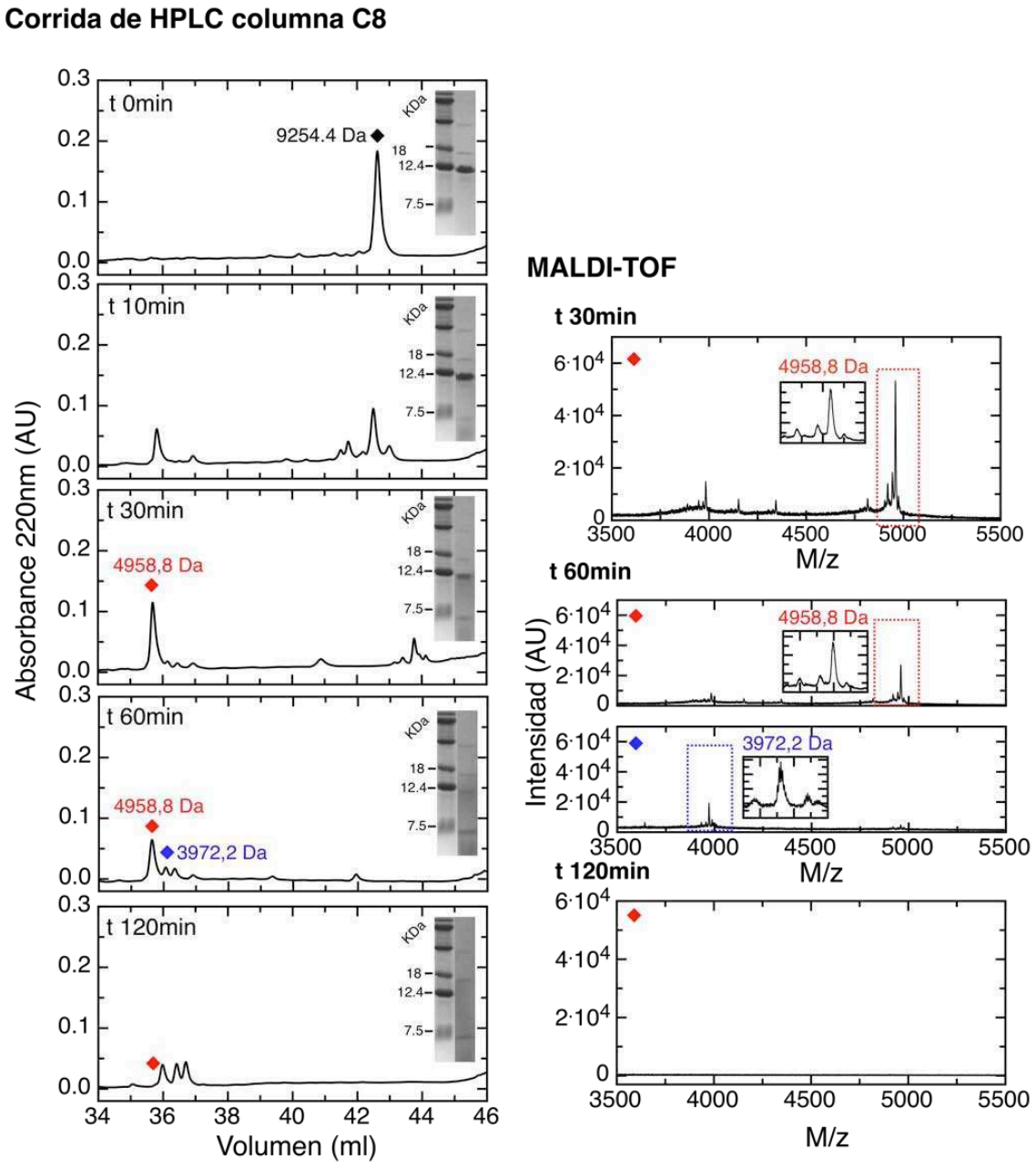
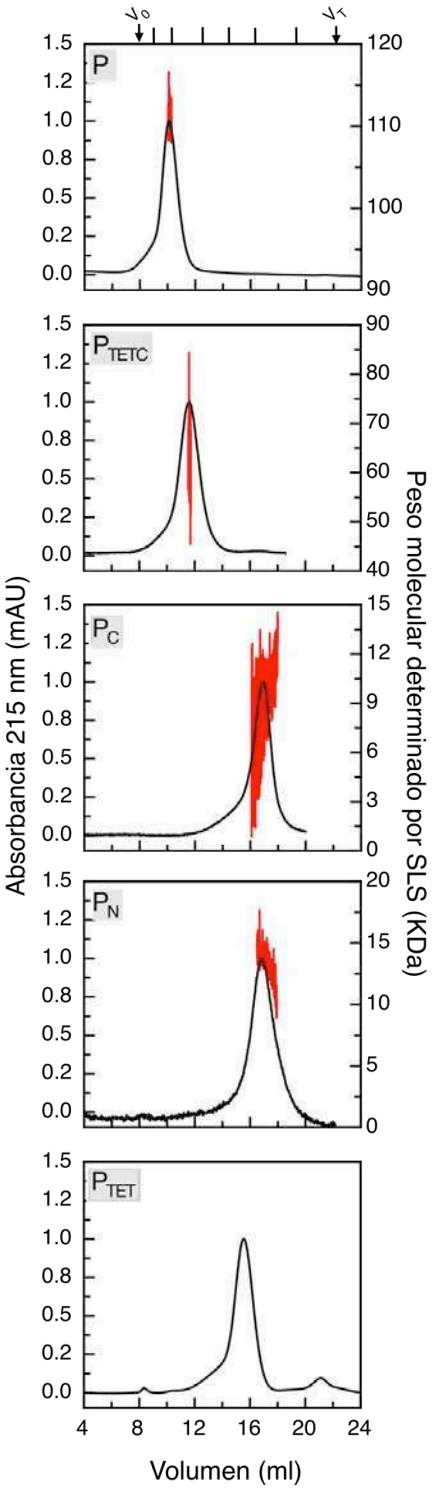


Figura A10. Cromatografía de exclusión molecular en superdex S200 analítica y determinación del peso molecular por SLS de los distintos fragmentos de P.



SLS 110.8 ± 1.9 KDa

SLS 62.7 ± 8.4 KDa

SLS 9.34 ± 2.5 KDa

SLS 13 ± 1.4 KDa

SLS: Esperante *et al.* 2012

Figura A11. Medidas de velocidad de sedimentación por AUC para E7N, CR1 y CR2. Superposición de las distribuciones *c(s)* a distintas concentraciones de péptidos. medidas en el rango de 0-2S. La señal fue normalizada a 1 cm de paso óptico. Todas las medidas de AUC fueron realizadas en 50 mM Tris-HCl (pH 7.5) y 1 mM DTT, 50.000 rpm y 20 ºC. 1 Svedberg= 10^{-13} segundos.

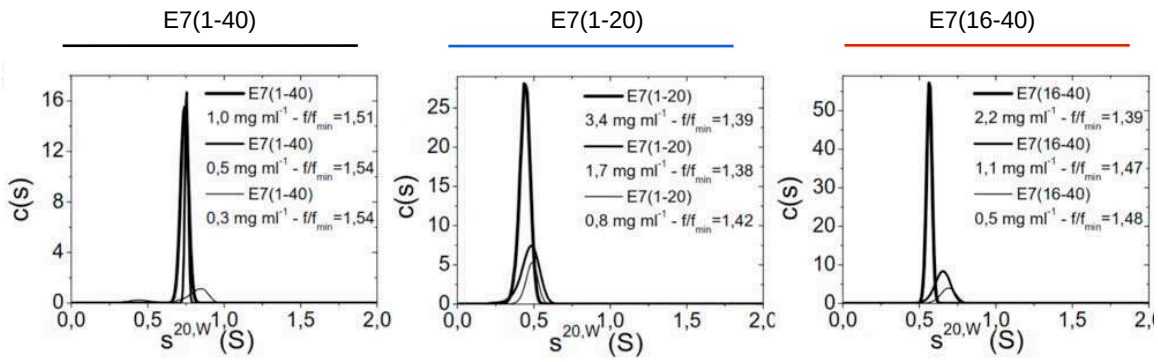


Figura A12. Análisis de los desplazamientos químicos del Cβ entre agua pH 7.5 y agua pH 5.0 medidas para E7N. Se analizaron los cambios de núcleo correspondiente al Cβ, el cual es el núcleo que más resiente cambios de pH, debido a que son los que están más cerca de las cadenas laterales de los grupos protonables. Todas las medidas fueron realizadas en agua pH 7.5 y pH 5.0, en 10 mM TCEP y 20 ºC. Los grupos cuyos desplazamientos químicos se ven más influenciados por el cambio de pH son las His y los residuos ácidos. Las Tyr y la Cys no son modificadas al cambiar el pH de 7.5 a 5.0 (barras azul).

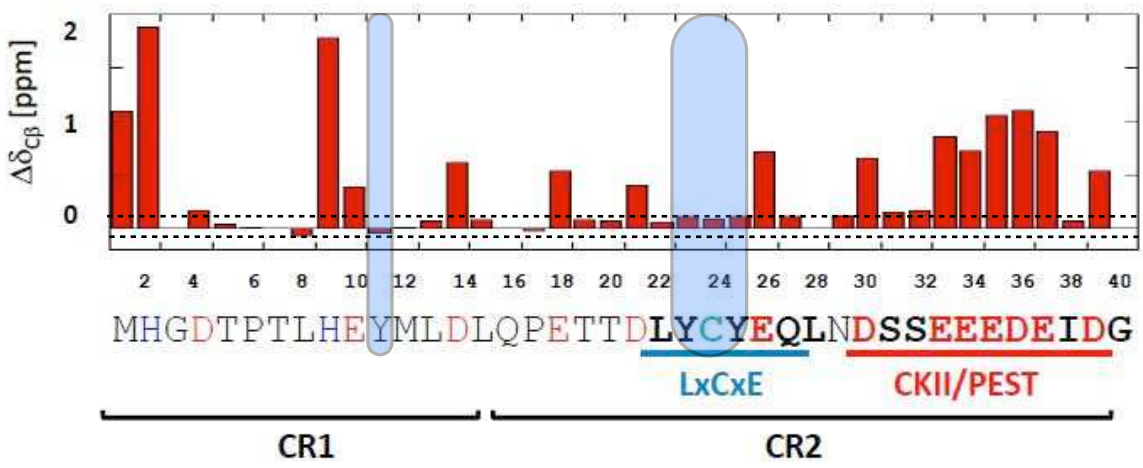


Figura A13. Alineamiento de secuencias de P de la subfamilia *Pneumovirinae* realizado con ClustalW. Género *Pneumovirus*: HRSVA2 (P03421), HRSV B1 (O42062), HRSV Long strain (P12579), ORSV (Q83956), BRSV (P33454), CanineRSV (W0ISD1), DogMPV (F1ACP3), MPV (Q5MKM7). Género *Metapneumovirus*: HMPV strain Can97 (Q8B9Q8), HMPV (Q6QQI3), HMPV (X4XZY6), AMPV_Can(Q2Y2M5), AMPV-C (A8Y7P9), AMPV-B (A8Y7P8), AMPV-D (A0A077SG92).

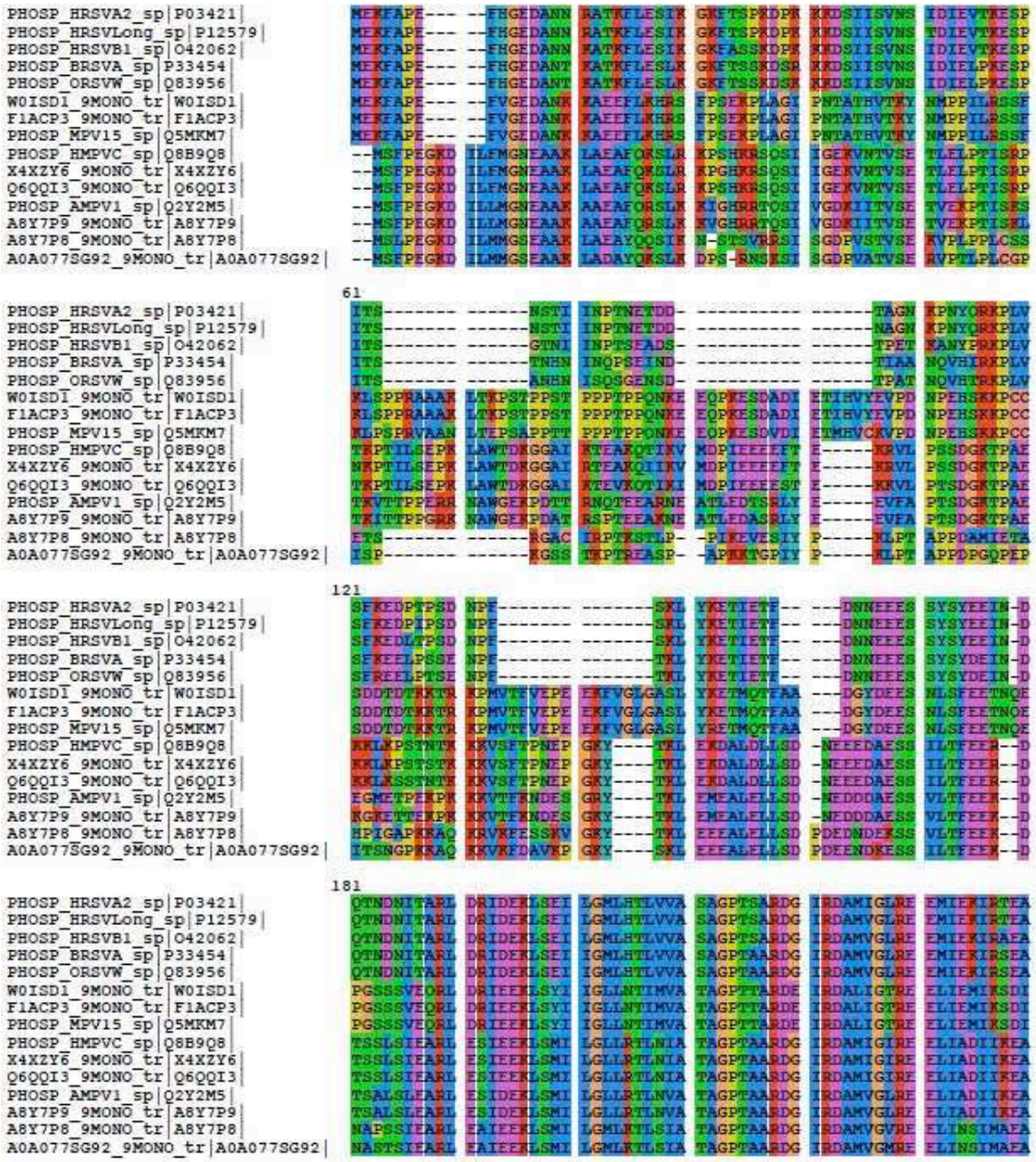


Figura A13. Alineamiento de secuencias de P de la subfamilia *Pneumovirinae* realizado con ClustalW. (Continuación)

```
                                          271
PHOSP_HRSVA2_sp|P03421|                   LMTNDRLEAM ARLRNEESEK MAKDTSDEVS LNPTSEKLNN LLEGN----- ---------DS
PHOSP_HRSVLong_sp|P12579|                 LMTNDRLEAM ARLRNEESEK MAKDTSDEVS LNPTSEKLNN LLEGN----- ---------DS
PHOSP_HRSVB1_sp|O42062|                   LMTNDRLEAM ARLRNEESEK MAKDTSDEVP LNPTSKKLSD LLEDN----- ---------DS
PHOSP_BRSVA_sp|P33454|                    LMTNDRLEAM ARLRDEESEK MTKDTSDEVR LTPTSEKLNM VLEDE----- ---------SS
PHOSP_ORSVW_sp|Q83956|                    LMTNDRLEAM ARLRNEESEK MAKDTSDDVN LNSTSEKLNT ILEED----- ---------NS
W0ISD1_9MONO_tr|W0ISD1|                   LTVNDRIAAM EKLRDEECSR AETDDGSACY LTDRARILDK IVSSNA---- --------EEA
F1ACP3_9MONO_tr|F1ACP3|                   LTVNDRVAAM EKLRDEECSR AETDDGSACY LTDRARILDK IVSSNA---- --------EEA
PHOSP_MPV15_sp|Q5MKM7|                    LTVNDRIVAM EKLRDEECSR ADTDDGSACY LTDRARILDK IVSSNA---- --------EEA
PHOSP_HMPVC_sp|Q8B9Q8|                    -----KGKAA EMMEEEMNQR TKIGNGS-VR LTEKAKELNK IVEDESTSGE SEEEEELKDT
X4XZY6_9MONO_tr|X4XZY6|                   -----KGKAA EMMEEEMNQR TKIGNGS-VR LTEKAKELNK IVEDESTSGE SEEEEEPKDT
Q6QQI3_9MONO_tr|Q6QQI3|                   -----KGKAA EMMEEEMNQR TKIGNGS-VR LTEKAKELNK IVEDESTSGE SEEEEELKDI
PHOSP_AMPV1_sp|Q2Y2M5|                    -----KGKAA EMMKEEAKQK SKIGNGS-VG LTEKAKELNK IVEDESTSGE SEEEEEEEDE
A8Y7P9_9MONO_tr|A8Y7P9|                   -----KGKAA EMMKEEAKQK SKIGNGS-VG LTEKAKELNK IVEDESTSGE SEEEEEEEDE
A8Y7P8_9MONO_tr|A8Y7P8|                   -----KGKIA EMIKEEDAQR AKIGDGS-VR LTEKAKELNR MLEDQSSSGE SETESEETEP
A0A077SG92_9MONO_tr|A0A077SG92|            -----KDKIS EMIKEEDTQR AKIGDGS-VR LTEKAKELNK VLEDQSSSGE SDSESEDEEP
```

```
                                          301
PHOSP_HRSVA2_sp|P03421|                   DNDLSLEDF- ----
PHOSP_HRSVLong_sp|P12579|                 DNDLSLEDF- ----
PHOSP_HRSVB1_sp|O42062|                   DNDLSLDDF- ----
PHOSP_BRSVA_sp|P33454|                    DNDLSLEDF- ----
PHOSP_ORSVW_sp|Q83956|                    DNDLSLEDF- ----
W0ISD1_9MONO_tr|W0ISD1|                   KEDLDVDDIM GINE
F1ACP3_9MONO_tr|F1ACP3|                   KEDLDVDDIM GINE
PHOSP_MPV15_sp|Q5MKM7|                    KEDLDVDDIM GINE
PHOSP_HMPVC_sp|Q8B9Q8|                    QENNQEDDIY QLIM
X4XZY6_9MONO_tr|X4XZY6|                   QENNQEDDIY QLIM
Q6QQI3_9MONO_tr|Q6QQI3|                   QENNQEDDIY QLIM
PHOSP_AMPV1_sp|Q2Y2M5|                    EESNPDDDLY SLTM
A8Y7P9_9MONO_tr|A8Y7P9|                   EESNPDDDLY SLTM
A8Y7P8_9MONO_tr|A8Y7P8|                   DTDGENDDIY SFDM
A0A077SG92_9MONO_tr|A0A077SG92|            EPD-ETADIY SLEL
```

Apéndice Tabla A1. Asignación de los desplazamientos químicos ^{1}H y ^{13}C en solución acuosa a pH 7.5 y 20 ℃ en 10 mM TCEP y 5% D_2O. Los desplazamientos químicos están reportados en ppm con una precisión de ±0.02 ppm. Los desplazamientos químicos del ^{13}C están reportados en ppm con una precisión de ±0.1 ppm. [b]Desplazamientos químicos del carbono primero y del proton en paréntesis.

Residuo	NH	$^{13}C_\alpha$	$^{1}H_\alpha$	$^{13}C_\beta$	$^{1}H_\beta$	otros[b]
M1		54.9	3.88	33.3	1.99	γ 30.6 (2.49/2.49), ε 16.7 (2.05)
H2		56.4	4.57	30.1	3.11/3.08	
G3		44.3	3.96/3.83			
D4	8.26	53.9	4.62	40.7	2.64/2.64	
T5	8.16	59.3	4.57	69.2	4.12	δ_2 20.8 (1.20)
P6		62.8	4.44	31.6	2.26/1.87	γ 26.8 (1.97/1.97), δ 50.5 (3.81/3.68)
T7	8.26	61.5	4.23	69.3	4.11	δ_2 21.1 (1.13)
L8	8.23	54.7	4.29	41.7	1.49/1.54	γ 26.4 (1.50), δ_1 23.0 (0.80), δ_2 24.2 (0.86)
H9	8.21	55.7	4.52	30.2	3.00/2.95	
E10	8.35	56.5	4.13	29.6	1.89/1.85	γ 35.7 (2.09/2.09)
Y11	8.15	57.4	4.53	38.0	3.03/2.94	
M12	8.15	54.9	4.37	32.2	2.02/1.92	γ 31.4 (2.46/2.41), ε 16.3 (2.03)
L13	8.02	54.8	4.24	41.9	1.58/1.54	γ 26.4 (1.57), δ_1 22.9 (0.83), δ_2 24.3 (0.89)
D14	8.24	53.7	4.55	40.4	2.66/2.74	
L15	8.00	54.4	4.30	41.9	1.58/1.54	γ 26.3 (1.55), δ_1 22.9 (0.82), δ_2 24.5 (0.88)
Q16	8.39	53.0	4.57	28.3	2.08/1.91	γ 32.9 (2.31/2.31)
P17		62.6	4.39	31.5	2.26/1.89	γ 26.8 (1.97/1.97), δ 50.0 (3.71/3.63)
E18	8.63	56.2	4.29	29.7	2.04/1.94	γ 35.8 (2.29/2.25)
T19	8.25	61.1	4.40	69.3	4.23	δ_2 21.0 (1.17)
T20	8.18	61.4	4.31	69.3	4.19	δ_2 20.9 (1.16)
D21	8.16	53.9	4.56	40.4	2.64/2.56	
L22	8.07	54.8	4.21	41.8	1.45/1.38	γ 26.3 (1.47), δ_1 22.9 (0.77), δ_2 24.3 (0.84)
Y23	8.15	57.4	4.46	38.0	2.92/2.87	
C24	7.99	57.7	4.38	27.3	2.75/2.75	
Y25	8.16	57.7	4.48	38.1	3.04/2.91	
E26	8.26	56.0	4.19	29.9	1.95/1.86	γ 35.8 (2.18/2.18)
Q27	8.29	55.1	4.26	28.9	2.05/1.95	γ 33.8 (2.31/2.31)
L28	8.31	54.5	4.32	41.8	1.61/1.54	γ 26.4 (1.58), δ_1 22.8 (0.82), δ_2 24.4 (0.88)
N29	8.45	52.6	4.71	38.7	2.79/2.69	
D30	8.37	53.7	4.61	40.7	2.69/2.65	
S31	8.32	57.7	4.48	63.3	3.89/3.85	
S32	8.46	58.0	4.45	63.4	3.87/3.85	
E33	8.47	56.0	4.30	29.8	2.05/1.89	γ 35.7 (2.24/2.24)
E34	8.39	55.9	4.26	29.7	2.00/1.88	γ 35.7 (2.24/2.24)
E35	8.44	55.9	4.27	30.1	1.89/2.02	γ 35.7 (2.25/2.22)
D36	8.41	53.9	4.55	40.9	2.64/2.56	
E37	8.39	55.9	4.26	29.8	2.00/1.89	γ 35.7 (2.25/2.20)
I38	8.25	60.8	4.22	38.2	1.83	γ_1 26.8 (1.42/1.16), δ_1 12.3 (0.83), γ_2 16.8 (0.86)
D39	8.53	54.1	4.53	40.4	2.71/2.64	
G40	8.39	44.7	3.87/3.87			

Apéndice Tabla A2. Asignación de los desplazamientos químicos ^{1}H y ^{13}C en solución acuosa a pH 7.5 y 20 ℃ en 10 mM TCEP y 50% TFE-d2. Los desplazamientos químicos están reportados en ppm con una precisión de ±0.02 ppm. Los desplazamientos químicos del ^{13}C están reportados en ppm con una precisión de ±0.1 ppm. [b]Desplazamientos químicos del carbono primero y del proton en paréntesis.

Residuo	NH	$^{13}C_{\alpha}$	$^{1}H_{\alpha}$	$^{13}C_{\beta}$	$^{1}H_{\beta}$	Otros[b]
M1		55.5	3.70	34.2	1.94/1.88	γ 30.9 (2.46/2.46), ε 15.7 (2.03
H2	7.77	56.2	4.53	30.2	3.12/3.04	
G3	8.25	44.7	3.98/3.85			
D4	8.10	53.8	4.68	40.7	2.70/2.70	
T5	7.88	59.5	4.59	69.4	4.17	δ_2 20.6 (1.28)
P6		63.1	4.47	31.3	2.23/1.87	γ 26.6 (1.95/2.00), δ 50.3 (3.78/3.68)
T7	7.96	61.3	4.34	70.0	4.26	δ_2 21.0 (1.21)
L8	7.86	56.2	4.25	41.7	1.58/1.58	γ 26.4 (1.58), δ_1 23.0 (0.83), δ_2 23.3 (0.88)
H9	7.68	57.7	4.35	30.2	3.03/2.98	
E10	8.02	58.5	3.94	29.0	1.99/1.94	γ 36.2 (2.16/2.16)
Y11	7.85	59.9	4.41	37.7	3.12/3.03	
M12	7.86	56.4	4.24	31.9	2.06/2.06	γ 31.8 (2.56/2.51), ε 15.8 (2.02)
L13	7.60	55.8	4.17	41.6	1.72/1.53	γ 26.4 (1.65), δ_1 22.5 (0.83), δ_2 24.1 (0.86)
D14	7.73	54.2	4.60	40.6	2.70/2.61	
L15	7.52	54.8	4.26	42.0	1.63/1.53	γ 26.3 (1.59), δ_1 22.4 (0.80), δ_2 24.1 (0.85)
Q16	7.85	53.2	4.60	28.5	2.16/1.98	γ 33.0 (2.37/2.37)
P17		63.3	4.38	31.4	2.25/1.90	γ 26.7 (2.00/1.95), δ 50.0 (3.71/3.71)
E18	8.71	56.9	4.29	29.4	2.08/1.98	γ 35.8 (2.32/2.32)
T19	7.93	62.2	4.30	69.0	4.30	δ_2 21.0 (1.21)
T20	7.79	62.9	4.22	69.0	4.26	δ_2 20.9 (1.21)
D21	8.02	55.1	4.55	40.3	2.68/2.68	
L22	7.71	56.2	4.18	41.5	1.60/1.53	γ 26.4 (1.53), δ_1 22.6 (0.80), δ_2 24.7 (0.88)
Y23	7.82	59.0	4.39	37.8	3.08/3.01	
C24	7.82	60.0	4.25	26.7	2.96/2.82	
Y25	7.95	59.9	4.24	37.7	3.02/3.02	
E26	8.25	58.0	4.04	29.1	2.03/2.03	γ 35.8 (2.28/2.18)
Q27	7.81	56.5	4.15	28.5	2.09/2.01	γ 33.3 (2.13/2.13)
L28	7.86	55.5	4.21	41.7	1.56/1.52	γ 26.3 (1.68), δ_1 22.2 (0.82), δ_2 24.1 (0.87)
N29	7.78	53.2	4.61	38.8	2.71/2.56	
D30	7.96	54.0	4.65	40.9	2.69/2.69	
S31	7.96	58.0	4.51	63.4	3.96/3.89	
S32	8.23	58.4	4.45	63.4	3.95/3.89	
E33	8.32	56.3	4.34	29.7	2.10/1.93	γ 35.7 (2.30/2.30)
E34	8.22	56.3	4.28	29.9	2.06/1.94	γ 35.7 (2.28/2.28)
E35	8.38	56.2	4.26	29.9	2.06/1.94	γ 35.7 (2.30/2.30)
D36	8.20	54.0	4.60	40.9	2.64/2.64	
E37	8.20	56.2	4.29	29.9	2.06/1.94	γ 35.8 (2.32/2.23)
I38	7.98	61.2	4.10	38.3	1.87	γ_1 26.9 (1.47/1.18), δ_1 12.1 (0.85), γ_2 16.6 (0.89)
D39	8.19	53.9	4.58	40.5	2.72/2.56	
G40	8.10	44.6	3.88/3.88			

Apéndice Tabla A3. Asignación de los desplazamientos químicos ^{1}H y ^{13}C en solución acuosa a pH 5.0 y 20 ℃ en 10 mM TCEP y 5% D_2O. Los desplazamientos químicos están reportados en ppm con una precisión de ±0.02 ppm. Los desplazamientos químicos del ^{13}C están reportados en ppm con una precisión de ±0.1 ppm. [b]Desplazamientos químicos del carbono primero y del proton en paréntesis.

Residuo	NH	$^{13}C_\alpha$	$^{1}H_\alpha$	$^{13}C_\beta$	$^{1}H_\beta$	Otros
M1		55.5	3.70	32.3	2.09	γ 30.3 (2.52/2.52), ε 16.2 (2.03)
H2	8.28	55.5	4.65	28.2	3.22/3.22	γ_2 120.1 (7.29), ε_1 136.5 (8.55)
G3	8.59	44.6	3.94/3.89			
D4	8.25	53.7	4.64	40.6	2.64/2.64	
T5	8.17	59.4	4.58	69.2	4.11	δ_2 20.8 (1.18)
P6		62.9	4.44	31.6	2.24/1.85	γ 26.8 (1.95/1.95), δ 50.5 (3.79/3.67)
T7	8.19	61.5	4.24	69.3	4.12	δ_2 21.1 (1.13)
L8	8.17	54.9	4.28	41.8	1.58/1.52	γ 26.4 (1.47), δ_1 23.1 (0.79), δ_2 24.1 (0.86)
H9	8.44	54.9	4.59	28.4	3.08/3.06	γ_2 119.4 (7.13), ε_1 135.7 (8.56)
E10	8.38	56.3	4.17	29.2	1.90/1.86	γ 34.9 (2.14/2.14)
Y11	8.18	57.6	4.52	38.0	3.01/2.92	$\gamma_{1,2}$ 132.5 (7.06), $\varepsilon_{1,2}$ 118.1 (6.75)
M12	8.16	55.0	4.36	32.1	2.00/1.91	γ 31.5 (2.46/2.40), ε 16.4 (2.01)
L13	7.99	54.9	4.24	41.8	1.58/1.52	γ 26.5 (1.53), δ_1 23.0 (0.81), δ_2 24.3 (0.86)
D14	8.22	53.5	4.57	40.4	2.71/2.58	
L15	7.97	54.5	4.29	41.8	1.57/1.53	γ 26.4 (1.57), δ_1 22.8 (0.79), δ_2 24.4 (0.86)
Q16	8.28	53.1	4.57	28.3	2.06/1.90	γ 33.0 (2.32/2.32), ($H\varepsilon_{21,22}$ 6.81/7.53)
P17		62.8	4.39	31.6	2.23/1.87	γ 26.9 (1.95/1.95), δ 50.0 (3.68/3.60)
E18	8.57	56.0	4.31	29.1	2.05/1.94	γ 34.6 (2.32/2.32)
T19	8.16	61.2	4.39	69.2	4.20	δ_2 21.0 (1.15)
T20	8.11	61.5	4.30	69.2	4.22	δ_2 20.9 (1.14)
D21	8.32	53.8	4.57	40.0	2.69/2.60	
L22	8.04	55.0	4.19	41.7	1.44/1.36	γ 26.3 (1.44), δ_1 22.9 (0.75), δ_2 24.3 (0.82)
Y23	8.08	57.6	4.46	37.9	2.92/2.86	$\gamma_{1,2}$ 133.0 (7.00), $\varepsilon_{1,2}$ 117.5 (6.74)
C24	7.98	58.0	4.36	27.2	2.75/2.75	
Y25	8.10	57.9	4.46	38.0	3.02/2.91	$\gamma_{1,2}$ 132.5 (7. 06), $\varepsilon_{1,2}$ 117.6 (6.76)
E26	8.18	55.9	4.20	29.2	1.96/1.87	γ 34.4 (2.23/2.23)
Q27	8.21	55.3	4.24	28.7	2.04/1.92	γ 33.3 (2.28/2.28), ($H\varepsilon_{1,2}$ 6.81/7.48)
L28	8.19	54.7	4.29	41.8	1.57/1.52	γ 26.4 (1.53), δ_1 22.8 (0.79), δ_2 24.4 (0.85)
N29	8.36	52.7	4.68	38.6	2.77/2.67	($H\delta_{21,22}$ 7.53/6.86)
D30	8.30	53.5	4.64	40.0	2.69/2.69	
S31	8.31	58.1	4.46	63.2	3.90/3.84	
S32	8.37	58.4	4.41	63.2	3.88/3.88	
E33	8.31	55.9	4.31	28.9	2.05/1.91	γ 34.2 (2.31/2.31)
E34	8.28	55.9	4.28	29.0	2.03/1.91	γ 34.0 (2.32/2.32)
E35	8.35	55.9	4.28	29.0	2.04/1.92	γ 34.0 (2.33/2.33)
D36	8.35	53.7	4.58	39.8	2.71/2.64	
E37	8.24	55.7	4.30	28.9	2.05/1.90	γ 34.0 (2.32/2.32)
I38	8.14	60.9	4.10	38.1	1.82	γ_1 26.8 (1.41/1.13), δ_1 11.9 (0.80), γ_2 16.8 (0.84)
D39	8.46	53.9	4.57	39.8	2.73/2.67	
G40	8.31	44.6	3.88/3.88			

Apéndice Tabla A4. Asignación de los desplazamientos químicos ^{1}H y ^{13}C en solución acuosa a pH 5.0 y 20 ℃ en 10 mM TCEP y 12% TFE-d2. Los desplazamientos químicos están reportados en ppm con una precisión de ±0.02 ppm. Los desplazamientos químicos del ^{13}C están reportados en ppm con una precisión de ±0.1 ppm. [b]Desplazamientos químicos del carbono primero y del proton en paréntesis.

Residuo	NH	$^{13}C_{\alpha}$	$^{1}H_{\alpha}$	$^{13}C_{\beta}$	$^{1}H_{\beta}$	Otros[b]
M1		54.5	4.11	32.3	2.09	γ 30.3 (2.52/2.52), ε 16.2 (2.03)
H2		55.6	4.57	28.3	3.21/3.21	
G3		44.5	3.90/3.83			
D4		53.7	4.58	40.6	2.65/2.65	
T5		59.3	4.53	69.2	4.14	δ_2 20.8 (1.18)
P6		63.0	4.38	31.5	2.24/1.85	γ 26.8 (1.95/1.95), δ 50.5 (3.79/3.69)
T7		61.4	4.20	69.4	4.10	δ_2 21.1 (1.15)
L8		55.2	4.20	41.7	1.58/1.52	γ 26.4 (1.49), δ_1 23.2 (0.80), δ_2 24.1 (0.86)
H9		55.3	4.48	28.4	3.11/3.08	
E10		56.8	4.05	29.1	1.90/1.86	γ 35.1 (2.16/2.16)
Y11		57.9	4.41	37.9	3.05/2.93	
M12		55.3	4.26	32.0	2.02/1.95	γ 31.5 (2.48/2.43), ε 16.4 (2.02)
L13		55.1	4.15	41.7	1.58/1.52	γ 26.5 (1.54), δ_1 29.9 (0.80), δ_2 24.3 (0.86)
D14		53.5	4.50	39.8	2.72/2.59	
L15		54.5	4.22	41.9	1.57/1.53	γ 26.4 (1.55), δ_1 22.8 (0.79), δ_2 24.5 (0.84)
Q16		53.0	4.51	28.3	2.091.91	γ 33.1 (2.32/2.32)
P17		62.9	4.32	31.5	2.23/1.87	γ 26.9 (1.95/1.95), δ 50.1 (3.68/3.61)
E18		56.2	4.24	29.1	2.05/1.94	γ 34.7 (2.32/2.32)
T19		61.3	4.32	69.1	4.20	δ_2 21.0 (1.15)
T20		61.8	4.22	69.2	4.24	δ_2 20.9 (1.14)
D21		53.8	4.58	39.9	2.70/2.62	
L22		55.2	4.13	41.7	1.44/1.36	γ 26.3 (1.46), δ_1 23.0 (0.75), δ_2 24.3 (0.81)
Y23		57.9	4.39	37.9	2.96/2.88	
C24		58.4	4.28	27.1	2.77/2.77	
Y25		58.1	4.40	38.0	2.95/2.95	
E26		56.3	4.11	29.1	1.96/1.87	γ 34.4 (2.25/2.25)
Q27		55.5	4.16	28.7	2.04/1.93	γ 33.4 (2.26/2.26)
L28		54.8	4.22	41.8	1.57/1.52	γ 26.4 (1.53), δ_1 22.8 (0.79), δ_2 24.2(0.86)
N29		52.7	4.61	38.6	2.75/2.65	
D30		53.5	4.58	40.1	2.69/2.69	
S31		58.2	4.36	63.2	3.91/3.86	
S32		58.4	4.35	63.2	3.88/3.88	
E33		56.0	4.24	29.0	2.05/1.91	γ 34.2 (2.31/2.31)
E34		56.0	4.20	29.9	2.03/1.91	γ 34.0 (2.32/2.32)
E35		56.0	4.20	29.0	2.04/1.92	γ 34.1 (2.33/2.33)
D36		53.8	4.51	39.8	2.72/2.65	
E37		55.8	4.22	28.9	3.22/3.22	γ 34.1 (2.32/2.32)
I38		60.9	4.03	38.2	1.84	γ_1 26.8 (1.43/1.15), δ_1 11.9 (0.80), γ_2 16.8 (0.84)
D39		53.7	4.58	40.0	2.64/2.64	
G40		44.6	3.84	3.80/3.80		

Apéndice Tabla A5. Constantes de acoplamiento escalar 1JCαHα medidas para E7N (Hz). La precisión de las medidas es de ±0.5 Hz. [b]TFE-d2. [c]No determinado.

Residuo	H2O 5.00, 20°C	H2O pH 7.50, 4°C	50% TFE-d2 pH 7.50, 20°C[b]
M1	149.9	145.6	147.7
H2	143.5	145.3	145.9
G3	-	n.d.	-
D4	145.2	143.5	144.9
T5	145.2	142.3	148.2
P6	148.5	149.0	148.6
T7	144.1	141.4	145.0
L8	n.d.[c]	143.9	146.5
H9	144.6	145.3	148.8
E10	144.3	143.1	148.8
Y11	146.8	143.8	152.1
M12	145.9	143.5	148.2
L13	n.d.	145.4	146.6
D14	145.6	142.8	143.9
L15	143.8	142.7	140.7
Q16	143.4	n.d.	n.d.
P17	147.8	148.6	153.6
E18	143.6	144.0	143.5
T19	140.8	144.0	141.7
T20	141.7	143.9	144.5
D21	n.d.	142.4	146.6
L22	143.2	142.6	147.8
Y23	145.8	143.8	153.7
C24	145.4	144.8	148.3
Y25	145.7	147.3	149.4
E26	n.d.	145.6	151.6
Q27	142.8	145.2	144.0
L28	n.d.	143.2	147.0
N29	144.6	143.2	n.d.
D30	142.1	143.2	145.7
S31	142.4	142.2	142.4
S32	142.7	142.4	142.8
E33	144.1	145.9	146.1
E34	n.d.	146.3	n.d.
E35	n.d.	n.d.	n.d.
D36	144.1	147.7	146.1
E37	n.d.	n.d.	145.9
I38	144.4	144.5	142.4
D39	n.d.	142.8	141.9
G40	-	-	-

Apéndice Tabla A6. Estimación del número de residuos en α-hélice para P y sus fragmentos.

	Número de residuos en α-hélice			Residuos estabilizados	Número de residuos en α-hélice*		Residuos estabilizados	Número de residuos en α-hélice**		Residuos estabilizados
	5 ℃	25 ℃	60 ℃	Δ(5-60 ℃)	0% TFE	50% TFE	Δ(50-0% TFE)	pH 7.5	pH3.0	Δ(pH3.0-pH7.5)
P	93	84	65	28	*nd*	*nd*	*nd*	*nd*	*nd*	*nd*
P_{TETC}	95	80	59	36	75	108	33	*nd*	*nd*	*nd*
P_{NTET}	56	56	55	1	55	79	24	*nd*	*nd*	*nd*
P_N	18	19	22	-4	19	46	27	15	17	2
P_C	26	19	16	10	19	62	43	18	32	14
P_{C*}	2	3	5	-3	3	10	7	3	4	1

*Estas medidas fueron realizadas a 25 °C.
**Estas medidas fueron realizadas a 20 °C.

REFERENCIAS

REFERENCIAS

1. Fischer, E., Einfluss der Configuration auf die Wirkung der Enzyme. Ber Dt Chem

Ges, 1894. **27**: p. 2985–2993.

2. Mirsky, A.E. and L. Pauling, On the Structure of Native, Denatured, and Coagulated Proteins. Proc Natl Acad Sci U S A, 1936. **22**(7): p. 439-47.

3. Kendrew, J.C., et al., A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature, 1958. **181**(4610): p. 662-6.

4. Perutz, M.F., et al., Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. Nature, 1960. **185**(4711): p. 416-22.

5. Mc, M.T., N.J. Hipp, and M.L. Groves, The separation of the components of alpha-casein. I. The preparation of alpha 1-casein. Arch Biochem Biophys, 1959. **83**(1): p. 35-43.

6. Boesch, C., et al., 1H nuclear-magnetic-resonance studies of the molecular conformation of monomeric glucagon in aqueous solution. Eur J Biochem, 1978. **91**(1): p. 209-14.

7. Kriwacki, R.W., et al., Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proc Natl Acad Sci U S A, 1996. **93**(21): p. 11504-9.

8. Penkett, C.J., et al., NMR analysis of main-chain conformational preferences in an unfolded fibronectin-binding protein. J Mol Biol, 1997. **274**(2): p. 152-9.

9. O'Neil, K.T., R.H. Hoess, and W.F. DeGrado, Design of DNA-binding peptides based on the leucine zipper motif. Science, 1990. **249**(4970): p. 774-8.

10. Bracken, C., et al., Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA. J Mol Biol, 1999. **285**(5): p. 2133-46.

11. Lee, S.H., et al., Understanding pre-structured motifs (PreSMos) in intrinsically unfolded proteins. Curr Protein Pept Sci, 2012. **13**(1): p. 34-54.

12. Romero, P., et al., Thousands of proteins likely to have long disordered regions. Pac Symp Biocomput, 1998: p. 437-48.

13. Schweitzer-Stenner, R., Conformational propensities and residual structures in unfolded peptides and proteins. Mol Biosyst, 2012. **8**(1): p. 122-33.

14. van der Lee, R., et al., Classification of intrinsically disordered regions and proteins. Chem Rev, 2014. **114**(13): p. 6589-631.

15. Uversky, V.N., Intrinsically disordered proteins from A to Z. Int J Biochem Cell Biol, 2011. **43**(8): p. 1090-103.

16. Habchi, J., et al., Introducing protein intrinsic disorder. Chem Rev, 2014. **114**(13): p. 6561-88.

17. Oldfield, C.J. and A.K. Dunker, Intrinsically disordered proteins and intrinsically disordered protein regions. Annu Rev Biochem, 2014. **83**: p. 553-84.

18. Pullen, R.A., et al., The relation of polypeptide hormone structure and flexibility to receptor binding: the relevance of X-ray studies on insulins, glucagon and human placental lactogen. Mol Cell Biochem, 1975. **8**(1): p. 5-20.

19. Cary, P.D., T. Moss, and E.M. Bradbury, High-resolution proton-magnetic-resonance studies of chromatin core particles. Eur J Biochem, 1978. **89**(2): p. 475-82.

20. Schweers, O., et al., Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. J Biol Chem, 1994. **269**(39): p. 24290-7.

21. Weinreb, P.H., et al., NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. Biochemistry, 1996. **35**(43): p. 13709-15.

22. Romero, P., et al., Sequence complexity of disordered protein. Proteins, 2001. **42**(1): p. 38-48.

23. Uversky, V.N., A protein-chameleon: conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders. J Biomol Struct Dyn, 2003. **21**(2): p. 211-34.

24. Livesay, D.R., Protein dynamics: dancing on an ever-changing free energy stage. Curr Opin Pharmacol, 2010. **10**(6): p. 706-8.

25. Uversky, V.N. and A.K. Dunker, *Understanding protein non-folding.* Biochim Biophys Acta, 2010. **1804**(6): p. 1231-64.

26. Uversky, V.N., J.R. Gillespie, and A.L. Fink, Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins, 2000. **41**(3): p. 415-27.

27. Uversky, V.N., Natively unfolded proteins: a point where biology waits for physics. Protein Sci, 2002. **11**(4): p. 739-56.

28. Demarest, S.J., et al., Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. Nature, 2002. **415**(6871): p. 549-53.

29. Kingston, R.L., et al., Structure of the nucleocapsid-binding domain from the mumps virus polymerase; an example of protein folding induced by crystallization. J Mol Biol, 2008. **379**(4): p. 719-31.

30. Chemes, L.B., et al., Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains. Methods Mol Biol, 2012. **895**: p. 387-404.

31. Uversky, V.N., Size-exclusion chromatography in structural analysis of intrinsically disordered proteins. Methods Mol Biol, 2012. **896**: p. 179-94.

32. Gast, K. and C. Fiedler, Dynamic and static light scattering of intrinsically disordered proteins. Methods Mol Biol, 2012. **896**: p. 137-61.

33. Salvay, A.G., G. Communie, and C. Ebel, Sedimentation velocity analytical ultracentrifugation for intrinsically disordered proteins. Methods Mol Biol, 2012. **896**: p. 91-105.

34. Wilkins, D.K., et al., Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry, 1999. **38**(50): p. 16424-31.

35. Mulder, F.A.A., M. Lundqvist, and R.M. Scheek, Nuclear Magnetic Resonance Spectroscopy Applied to (Intrinsically) Disordered Proteins, in Instrumental Analysis of Intrinsically Disordered Proteins. 2010, John Wiley & Sons, Inc. p. 59-87.

36. Theillet, F.X., et al., Structural disorder of monomeric alpha-synuclein persists in mammalian cells. Nature, 2016. **530**(7588): p. 45-50.

37. Daughdrill, G.W., Determining Structural Ensembles for Intrinsically Disordered Proteins, in Instrumental Analysis of Intrinsically Disordered Proteins. 2010, John Wiley & Sons, Inc. p. 107-129.

38. Schuler, B. and W.A. Eaton, Protein folding studied by single-molecule FRET. Curr Opin Struct Biol, 2008. **18**(1): p. 16-26.

39. Shi, Z., et al., Conformation of the backbone in unfolded proteins. Chem Rev, 2006. **106**(5): p. 1877-97.

40. Adzhubei, A.A., M.J. Sternberg, and A.A. Makarov, Polyproline-II helix in proteins: structure and function. J Mol Biol, 2013. **425**(12): p. 2100-32.

41. Garcia-Alai, M.M., et al., Molecular basis for phosphorylation-dependent, PEST-mediated protein turnover. Structure, 2006. **14**(2): p. 309-19.

42. Fuxreiter, M., P. Tompa, and I. Simon, Local structural disorder imparts plasticity on linear motifs. Bioinformatics, 2007. **23**(8): p. 950-6.

43. Dinkel, H., et al., The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res, 2014. **42**(Database issue): p. D259-66.

44. Mohan, A., et al., Analysis of molecular recognition features (MoRFs). J Mol Biol, 2006. **362**(5): p. 1043-59.

45. Oldfield, C.J., et al., Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genomics, 2008. **9 Suppl 1**: p. S1.

46. Berlow, R.B., H.J. Dyson, and P.E. Wright, Functional advantages of dynamic protein disorder. FEBS Lett, 2015. **589**(19 Pt A): p. 2433-40.

47. Tompa, P. and M. Fuxreiter, Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem Sci, 2008. **33**(1): p. 2-8.

48. Fuxreiter, M. and P. Tompa, Fuzzy complexes: a more stochastic view of protein function. Adv Exp Med Biol, 2012. **725**: p. 1-14.

49. Xue, B., A.K. Dunker, and V.N. Uversky, Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. J Biomol Struct Dyn, 2012. **30**(2): p. 137-49.

50. Ward, J.J., et al., Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol, 2004. **337**(3): p. 635-45.

51. Schad, E., P. Tompa, and H. Hegyi, The relationship between proteome size, structural disorder and organism complexity. Genome Biol, 2011. **12**(12): p. R120.

52. Xue, B., et al., Structural disorder in viral proteins. Chem Rev, 2014. **114**(13): p. 6880-911.

53. Pushker, R., et al., Marked variability in the extent of protein disorder within and between viral families. PLoS One, 2013. **8**(4): p. e60724.

54. Drake, J.W., et al., Rates of spontaneous mutation. Genetics, 1998. **148**(4): p. 1667-86.

55. Davey, N.E., G. Trave, and T.J. Gibson, *How viruses hijack cell regulation.* Trends Biochem Sci, 2011. **36**(3): p. 159-69.

56. Bychkovsky, B.L., et al., Cervical cancer control in Latin America: A call to action. Cancer, 2016. **122**(4): p. 502-14.

57. Capote Negrin, L.G., Epidemiology of cervical cancer in Latin America. Ecancermedicalscience, 2015. **9**: p. 577.

58. Moody, C.A. and L.A. Laimins, Human papillomavirus oncoproteins: pathways to transformation. Nat Rev Cancer, 2010. **10**(8): p. 550-60.

59. Selvey, L.A., et al., Human papillomavirus (HPV) type 18 E7 protein is a short-lived steroid-inducible phosphoprotein in HPV-transformed cell lines. J Gen Virol, 1994. **75 (Pt 7)**: p. 1647-53.

60. Barbosa, M.S. and R. Schlegel, The E6 and E7 genes of HPV-18 are sufficient for inducing two-stage in vitro transformation of human keratinocytes. Oncogene, 1989. **4**(12): p. 1529-32.

61. Hebner, C.M. and L.A. Laimins, Human papillomaviruses: basic mechanisms of pathogenesis and oncogenicity. Rev Med Virol, 2006. **16**(2): p. 83-97.

62. Alonso, L.G., et al., High-risk (HPV16) human papillomavirus E7 oncoprotein is highly stable and extended, with conformational transitions that could explain its multiple cellular binding partners. Biochemistry, 2002. **41**(33): p. 10510-8.

63. Uversky, V.N., et al., Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. J Proteome Res, 2006. **5**(8): p. 1829-42.

64. Garcia-Alai, M.M., L.G. Alonso, and G. de Prat-Gay, The N-terminal module of HPV16 E7 is an intrinsically disordered domain that confers conformational and recognition plasticity to the oncoprotein. Biochemistry, 2007. **46**(37): p. 10405-12.

65. Chemes, L.B., et al., Sequence evolution of the intrinsically disordered and globular domains of a model viral oncoprotein. PLoS One, 2012. **7**(10): p. e47661.

66. Fassolari, M., et al., Minute time scale prolyl isomerization governs antibody recognition of an intrinsically disordered immunodominant epitope. J Biol Chem, 2013. **288**(18): p. 13110-23.

67. Smal, C., et al., Ordered self-assembly mechanism of a spherical oncoprotein oligomer triggered by zinc removal and stabilized by an intrinsically disordered domain. PLoS One, 2012. **7**(5): p. e36457.

68. Alonso, L.G., et al., Chaperone holdase activity of human papillomavirus E7 oncoprotein. Biochemistry, 2006. **45**(3): p. 657-67.

69. Dantur, K., et al., Cytosolic accumulation of HPV16 E7 oligomers supports different transformation routes for the prototypic viral oncoprotein: the amyloid-cancer connection. Int J Cancer, 2009. **125**(8): p. 1902-11.

70. Chemes, L.B., et al., Cysteine-rich positions outside the structural zinc motif of human papillomavirus E7 provide conformational modulation and suggest functional redox roles. Biochemistry, 2014. **53**(10): p. 1680-96.

71. Lay, M.K., et al., Advances in understanding respiratory syncytial virus infection in airway epithelial cells and consequential effects on the immune response. Microbes Infect, 2013. **15**(3): p. 230-42.

72. Hall, C.B., et al., The burden of respiratory syncytial virus infection in young children. N Engl J Med, 2009. **360**(6): p. 588-98.

73. Nair, H., et al., Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. Lancet, 2010. **375**(9725): p. 1545-55.

74. Collins, P.L. and J.A. Melero, Progress in understanding and controlling respiratory syncytial virus: still crazy after all these years. Virus Res, 2011. **162**(1-2): p. 80-99.

75. Marcone, D.N., et al., Viral etiology of acute respiratory infections in hospitalized and outpatient children in Buenos Aires, Argentina. Pediatr Infect Dis J, 2013. **32**(3): p. e105-10.

76. Krilov, L.R., Respiratory syncytial virus disease: update on treatment and prevention. Expert Rev Anti Infect Ther, 2011. **9**(1): p. 27-32.

77. Collins, P.L., *Human Respiratory Syncytial Virus*, in *The Biology of Paramyxoviruses*, S.K. Samal, Editor. 2011, Caister Academic Press.

78. Ortin, J. and J. Martin-Benito, The RNA synthesis machinery of negative-stranded RNA viruses. Virology, 2015. **479-480**: p. 532-44.

79. Cox, R. and R.K. Plemper, The paramyxovirus polymerase complex as a target for next-generation anti-paramyxovirus therapeutics. Front Microbiol, 2015. **6**: p. 459.

80. Tawar, R.G., et al., Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. Science, 2009. **326**(5957): p. 1279-83.

81. Blondot, M.L., et al., Structure and functional analysis of the RNA- and viral phosphoprotein-binding domain of respiratory syncytial virus M2-1 protein. PLoS Pathog, 2012. **8**(5): p. e1002734.

82. Tanner, S.J., et al., Crystal structure of the essential transcription antiterminator M2-1 protein of human respiratory syncytial virus and implications of its phosphorylation. Proc Natl Acad Sci U S A, 2014. **111**(4): p. 1580-5.

83. Habchi, J. and S. Longhi, Structural Disorder within Paramyxoviral Nucleoproteins and Phosphoproteins in Their Free and Bound Forms: From Predictions to Experimental Assessment. Int J Mol Sci, 2015. **16**(7): p. 15688-726.

84. Yabukarski, F., et al., Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone. Nat Struct Mol Biol, 2014. **21**(9): p. 754-9.

85. Leyrat, C., et al., Structure of the vesicular stomatitis virus N(0)-P complex. PLoS Pathog, 2011. **7**(9): p. e1002248.

86. Llorente, M.T., et al., Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an alpha-helical domain involved in oligomerization. J Gen Virol, 2006. **87**(Pt 1): p. 159-69.

87. Llorente, M.T., et al., Structural properties of the human respiratory syncytial virus P protein: evidence for an elongated homotetrameric molecule that is the smallest orthologue within the family of paramyxovirus polymerase cofactors. Proteins, 2008. **72**(3): p. 946-58.

88. Asenjo, A., et al., Residues in human respiratory syncytial virus P protein that are essential for its activity on RNA viral synthesis. Virus Res, 2008. **132**(1-2): p. 160-73.

89. Asenjo, A., L. Rodriguez, and N. Villanueva, Determination of phosphorylated residues from human respiratory syncytial virus P protein that are dynamically dephosphorylated by cellular phosphatases: a possible role for serine 54. J Gen Virol, 2005. **86**(Pt 4): p. 1109-20.

90. Asenjo, A., E. Calvo, and N. Villanueva, Phosphorylation of human respiratory syncytial virus P protein at threonine 108 controls its interaction with the M2-1 protein in the viral RNA polymerase complex. J Gen Virol, 2006. **87**(Pt 12): p. 3637-42.

91. Sanchez-Seco, M.P., et al., C-terminal phosphorylation of human respiratory syncytial virus P protein occurs mainly at serine residue 232. J Gen Virol, 1995. **76 (Pt 2)**: p. 425-30.

92. Mazumder, B., G. Adhikary, and S. Barik, Bacterial expression of human respiratory syncytial viral phosphoprotein P and identification of Ser237 as the site of phosphorylation by cellular casein kinase II. Virology, 1994. **205**(1): p. 93-103.

93. Navarro, J., C. Lopez-Otin, and N. Villanueva, Location of phosphorylated residues in human respiratory syncytial virus phosphoprotein. J Gen Virol, 1991. **72 (Pt 6)**: p. 1455-9.

94. Asenjo, A., J.C. Gonzalez-Armas, and N. Villanueva, Phosphorylation of human respiratory syncytial virus P protein at serine 54 regulates viral uncoating. Virology, 2008. **380**(1): p. 26-33.

95. Dupuy, L.C., et al., Casein kinase 2-mediated phosphorylation of respiratory syncytial virus phosphoprotein P is essential for the transcription elongation activity of the viral polymerase; phosphorylation by casein kinase 1 occurs mainly at Ser(215) and is without effect. J Virol, 1999. **73**(10): p. 8384-92.

96. Galloux, M., et al., Identification and characterization of the binding site of the respiratory syncytial virus phosphoprotein to RNA-free nucleoprotein. J Virol, 2015. **89**(7): p. 3484-96.

97. Sourimant, J., et al., Fine mapping and characterization of the L-polymerase-binding domain of the respiratory syncytial virus phosphoprotein. J Virol, 2015. **89**(8): p. 4421-33.

98. Tran, T.L., et al., The nine C-terminal amino acids of the respiratory syncytial virus protein P are necessary and sufficient for binding to ribonucleoprotein complexes in which six ribonucleotides are contacted per N protein protomer. J Gen Virol, 2007. **88**(Pt 1): p. 196-206.

99. Ouizougun-Oubari, M., et al., A Druggable Pocket at the Nucleocapsid/Phosphoprotein Interaction Site of Human Respiratory Syncytial Virus. J Virol, 2015. **89**(21): p. 11129-43.

100. Slack, M.S. and A.J. Easton, Characterization of the interaction of the human respiratory syncytial virus phosphoprotein and nucleocapsid protein using the two-hybrid system. Virus Res, 1998. **55**(2): p. 167-76.

101. Khattar, S.K., A.S. Yunus, and S.K. Samal, Mapping the domains on the phosphoprotein of bovine respiratory syncytial virus required for N-P and P-L interactions using a minigenome system. J Gen Virol, 2001. **82**(Pt 4): p. 775-9.

102. Lu, B., et al., Identification of temperature-sensitive mutations in the phosphoprotein of respiratory syncytial virus that are likely involved in its interaction with the nucleoprotein. J Virol, 2002. **76**(6): p. 2871-80.

103. Tcherkasskaya, O., E.A. Davidson, and V.N. Uversky, *Biophysical constraints for protein structure prediction.* J Proteome Res, 2003. **2**(1): p. 37-42.

104. Soler-Gonzalez, A.S. and A.R. Fersht, *Helix stability in barstar peptides.* Eur J Biochem, 1997. **249**(3): p. 724-32.

105. Munoz, V. and L. Serrano, Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. J Mol Biol, 1995. **245**(3): p. 297-308.

106. Liu, X. and R. Marmorstein, Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor. Genes Dev, 2007. **21**(21): p. 2711-6.

107. Lehrman, S.R., J.L. Tuls, and M. Lund, Peptide alpha-helicity in aqueous trifluoroethanol: correlations with predicted alpha-helicity and the secondary structure of the corresponding regions of bovine growth hormone. Biochemistry, 1990. **29**(23): p. 5590-6.

108. Jasanoff, A. and A.R. Fersht, Quantitative determination of helical propensities from trifluoroethanol titration curves. Biochemistry, 1994. **33**(8): p. 2129-35.

109. Pace, C.N., G.R. Grimsley, and J.M. Scholtz, Protein ionizable groups: pK values and their contribution to protein stability and solubility. J Biol Chem, 2009. **284**(20): p. 13285-9.

110. Zhong, L. and W.C. Johnson, Jr., Environment affects amino acid preference for secondary structure. Proc Natl Acad Sci U S A, 1992. **89**(10): p. 4462-5.

111. Wu, C.S., K. Ikeda, and J.T. Yang, Ordered conformation of polypeptides and proteins in acidic dodecyl sulfate solution. Biochemistry, 1981. **20**(3): p. 566-70.

112. Shen, Y. and A. Bax, Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. J Biomol NMR, 2010. **46**(3): p. 199-204.

113. Kashtanov, S., et al., Using chemical shifts to assess transient secondary structure and generate ensemble structures of intrinsically disordered proteins. Methods Mol Biol, 2012. **895**: p. 139-52.

114. Eliezer, D., Biophysical characterization of intrinsically disordered proteins. Curr Opin Struct Biol, 2009. **19**(1): p. 23-30.

115. Wishart, D.S. and B.D. Sykes, Chemical shifts as a tool for structure determination. Methods Enzymol, 1994. **239**: p. 363-92.

116. Shi, Z., R.W. Woody, and N.R. Kallenbach, Is polyproline II a major backbone conformation in unfolded proteins? Adv Protein Chem, 2002. **62**: p. 163-240.

117. Vuister, G.W., F. Delaglio, and A. Bax, The use of 1JC alpha H alpha coupling constants as a probe for protein backbone conformation. J Biomol NMR, 1993. **3**(1): p. 67-80.

118. Smith, M.D. and M. Jelokhani-Niaraki, pH-induced changes in intrinsically disordered proteins. Methods Mol Biol, 2012. **896**: p. 223-31.

119. Mikhonin, A.V., et al., UV resonance Raman determination of polyproline II, extended 2.5(1)-helix, and beta-sheet Psi angle energy landscape in poly-L-lysine and poly-L-glutamic acid. J Am Chem Soc, 2005. **127**(21): p. 7712-20.

120. Kim, M.K. and Y.K. Kang, Positional preference of proline in alpha-helices. Protein Sci, 1999. **8**(7): p. 1492-9.

121. Lee, C., et al., Contribution of proline to the pre-structuring tendency of transient helical secondary structure elements in intrinsically disordered proteins. Biochim Biophys Acta, 2014. **1840**(3): p. 993-1003.

122. Schneider, T.D. and R.M. Stephens, Sequence logos: a new way to display consensus sequences. Nucleic Acids Res, 1990. **18**(20): p. 6097-100.

123. Borcherds, W., et al., Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. Nat Chem Biol, 2014. **10**(12): p. 1000-2.

124. Chemes, L.B., G. de Prat-Gay, and I.E. Sanchez, Convergent evolution and mimicry of protein linear motifs in host-pathogen interactions. Curr Opin Struct Biol, 2015. **32**: p. 91-101.

125. Lee, C., et al., Structural basis for the recognition of the E2F transactivation domain by the retinoblastoma tumor suppressor. Genes Dev, 2002. **16**(24): p. 3199-212.

126. Fuxreiter, M., et al., Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol, 2004. **338**(5): p. 1015-26.

127. Lee, J.O., A.A. Russo, and N.P. Pavletich, Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. Nature, 1998. **391**(6670): p. 859-65.

128. Uversky, V.N., Intrinsically disordered proteins may escape unwanted interactions via functional misfolding. Biochim Biophys Acta, 2011. **1814**(5): p. 693-712.

129. Jansma, A.L., et al., The high-risk HPV16 E7 oncoprotein mediates interaction between the transcriptional coactivator CBP and the retinoblastoma protein pRb. J Mol Biol, 2014. **426**(24): p. 4030-48.

130. Chemes, L.B., et al., Evolution of linear motifs within the papillomavirus E7 oncoprotein. J Mol Biol, 2012. **422**(3): p. 336-46.

131. Chemes, L.B., et al., Targeting mechanism of the retinoblastoma tumor suppressor by a prototypical viral oncoprotein. Structural modularity, intrinsic disorder and phosphorylation of human papillomavirus E7. FEBS J, 2010. **277**(4): p. 973-88.

132. Esperante, S.A., G. Paris, and G. de Prat-Gay, Modular unfolding and dissociation of the human respiratory syncytial virus phosphoprotein p and its interaction with the m(2-1) antiterminator: a singular tetramer-tetramer interface arrangement. Biochemistry, 2012. **51**(41): p. 8100-10.

133. Hawe, A., M. Sutter, and W. Jiskoot, Extrinsic fluorescent dyes as tools for protein characterization. Pharm Res, 2008. **25**(7): p. 1487-99.

134. Lavery, D.N. and I.J. McEwan, Structural characterization of the native NH2-terminal transactivation domain of the human androgen receptor: a collapsed disordered conformation underlies structural plasticity and protein-induced folding. Biochemistry, 2008. **47**(11): p. 3360-9.

135. Bailey, R.W., et al., Clusterin, a binding protein with a molten globule-like region. Biochemistry, 2001. **40**(39): p. 11828-40.

136. Cox, R., et al., Structural studies on the authentic mumps virus nucleocapsid showing uncoiling by the phosphoprotein. Proc Natl Acad Sci U S A, 2014. **111**(42): p. 15208-13.

137. Shu, Y., et al., Plasticity in structural and functional interactions between the phosphoprotein and nucleoprotein of measles virus. J Biol Chem, 2012. **287**(15): p. 11951-67.

138. Habchi, J. and S. Longhi, Structural disorder within paramyxovirus nucleoproteins and phosphoproteins. Mol Biosyst, 2012. **8**(1): p. 69-81.

139. Habchi, J., et al., Structural disorder within Henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment. PLoS One, 2010. **5**(7): p. e11684.

140. Perez-Bercoff, A., J. Koch, and T.R. Burglin, LogoBar: bar graph visualization of protein logos with gaps. Bioinformatics, 2006. **22**(1): p. 112-4.

141. Leyrat, C., et al., Solution and crystallographic structures of the central region of the phosphoprotein from human metapneumovirus. PLoS One, 2013. **8**(11): p. e80371.

142. Karlin, D. and R. Belshaw, Detecting remote sequence homology in disordered proteins: discovery of conserved motifs in the N-termini of Mononegavirales phosphoproteins. PLoS One, 2012. **7**(3): p. e31719.

143. Communie, G., et al., Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus. PLoS Pathog, 2013. **9**(9): p. e1003631.

144. Gely, S., et al., Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein. J Mol Recognit, 2010. **23**(5): p. 435-47.

145. Gruet, A., et al., Fuzzy regions in an intrinsically disordered protein impair protein-protein interactions. FEBS J, 2016. **283**(4): p. 576-94.

146. Dyson, H.J. and P.E. Wright, Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol, 2005. **6**(3): p. 197-208.

147. Thurlkill, R.L., et al., pK values of the ionizable groups of proteins. Protein Sci, 2006. **15**(5): p. 1214-8.

148. Chen, Y.H., J.T. Yang, and K.H. Chau, Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. Biochemistry, 1974. **13**(16): p. 3350-9.

149. Greenfield, N.J., Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat Protoc, 2006. **1**(6): p. 2527-35.

150. Kyte, J. and R.F. Doolittle, A simple method for displaying the hydropathic character of a protein. J Mol Biol, 1982. **157**(1): p. 105-32.

151. Antonsson, A., et al., The human papillomavirus type 16 E7 protein binds human interferon regulatory factor-9 via a novel PEST domain required for transformation. J Interferon Cytokine Res, 2006. **26**(7): p. 455-61.

152. Wu, L., et al., E2F-Rb complexes assemble and inhibit cdc25A transcription in cervical carcinoma cells following repression of human papillomavirus oncogene expression. Mol Cell Biol, 2000. **20**(19): p. 7059-67.

153. Davies, R., et al., Human papillomavirus type 16 E7 associates with a histone H1 kinase and with p107 through sequences necessary for transformation. J Virol, 1993. **67**(5): p. 2521-8.

154. Oh, K.J., et al., The papillomavirus E7 oncoprotein is ubiquitinated by UbcH7 and Cullin 1- and Skp2-containing E3 ligase. J Virol, 2004. **78**(10): p. 5338-46.

155. Huh, K.W., et al., Association of the human papillomavirus type 16 E7 oncoprotein with the 600-kDa retinoblastoma protein-associated factor, p600. Proc Natl Acad Sci U S A, 2005. **102**(32): p. 11492-7.

156. Avvakumov, N., J. Torchia, and J.S. Mymryk, Interaction of the HPV E7 proteins with the pCAF acetyltransferase. Oncogene, 2003. **22**(25): p. 3833-41.

157. Campo-Fernandez, B., et al., Identification of the FHL2 transcriptional coactivator as a new functional target of the E7 oncoprotein of human papillomavirus type 16. J Virol, 2007. **81**(2): p. 1027-32.

158. Perea, S.E., P. Massimi, and L. Banks, Human papillomavirus type 16 E7 impairs the activation of the interferon regulatory factor-1. Int J Mol Med, 2000. **5**(6): p. 661-6.

159. Benevolo, M., et al., Comparative evaluation of nm23 and p16 expression as biomarkers of high-risk human papillomavirus infection and cervical intraepithelial neoplasia 2(+) lesions of the uterine cervix. Histopathology, 2010. **57**(4): p. 580-6.

160. Rey, O., et al., The E7 oncoprotein of human papillomavirus type 16 interacts with F-actin in vitro and in vivo. Virology, 2000. **268**(2): p. 372-81.

161. Chakrabarti, O. and S. Krishna, Molecular interactions of 'high risk' human papillomaviruses E6 and E7 oncoproteins: implications for tumour progression. J Biosci, 2003. **28**(3): p. 337-48.

162. Firzlaff, J.M., et al., The E7 protein of human papillomavirus type 16 is phosphorylated by casein kinase II. New Biol, 1989. **1**(1): p. 44-53.

163. Angeline, M., E. Merle, and J. Moroianu, The E7 oncoprotein of high-risk human papillomavirus type 16 enters the nucleus via a nonclassical Ran-dependent pathway. Virology, 2003. **317**(1): p. 13-23.

164. Lee, D.K., et al., The human papilloma virus E7 oncoprotein inhibits transforming growth factor-beta signaling by blocking binding of the Smad complex to its target sequence. J Biol Chem, 2002. **277**(41): p. 38557-64.

165. Dyson, N., et al., Homologous sequences in adenovirus E1A and human papillomavirus E7 proteins mediate interaction with the same set of cellular proteins. J Virol, 1992. **66**(12): p. 6893-902.

166. Martin, O.A., et al., Analysis of 13Calpha and 13Cbeta chemical shifts of cysteine and cystine residues in proteins: a quantum chemical approach. J Biomol NMR, 2010. **46**(3): p. 217-25.

167. Eccles, R., An explanation for the seasonality of acute upper respiratory tract viral infections. Acta Otolaryngol, 2002. **122**(2): p. 183-91.