

Tesis Doctoral

E2, regulador de transcripción y replicación del papilomavirus humano: interacción con ADN, dinámica conformacional y divergencia evolutiva

Dellarole, Mariano

2010

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:

Dellarole, Mariano. (2010). E2, regulador de transcripción y replicación del papilomavirus humano: interacción con ADN, dinámica conformacional y divergencia evolutiva. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Dellarole, Mariano. "E2, regulador de transcripción y replicación del papilomavirus humano: interacción con ADN, dinámica conformacional y divergencia evolutiva". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2010.

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires



UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE QUIMICA BIOLOGICA

“E2, regulador de transcripción y replicación del
papilomavirus humano: interacción con ADN, dinámica
conformacional y divergencia evolutiva”

Lic. Mariano Dellarole

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

Director de Tesis: Dr. Gonzalo de Prat Gay

Consejero de Estudios: Dr. Eduardo Tomás Cánepa

Lugar de trabajo: Laboratorio de Estructura-Función e Ingeniería de Proteínas, Fundación
Instituto Leloir e Instituto de Investigaciones Bioquímicas de Buenos Aires - Consejo
Nacional de Investigaciones Científicas y Técnicas, Patricias Argentinas 435, Buenos Aires,
Argentina.

Ciudad Autónoma de Buenos Aires, Argentina, seis de Diciembre de 2010

Resumen

“E2, regulador de transcripción y replicación del papilomavirus humano: interacción con ADN, dinámica conformacional y divergencia evolutiva”

Más de 50 tipos de papilomavirus humanos (HPV) infectan epitelio mucoso causando una variedad de lesiones benignas y malignas. El dominio C-terminal de la proteína E2 de HPV (E2C) discrimina entre cuatro sitios de ADN altamente similares regulando el ciclo de vida del virus vía la activación de la replicación y la activación y/o represión de la transcripción. Con el fin de tratar de entender la naturaleza de la interacción entre E2C y ADN, realizamos un análisis detallado del complejo utilizando herramientas bioinformáticas y el contenido de información en las secuencias proteicas y nucleotídicas y usando técnicas biofísicas, proteínas E2C recombinantes y sitios sintéticos de ADN específicos y no-específicos. Antes de estudiar el rol de la interacción E2C-ADN, nos focalizamos en el proceso de polimerización no-funcional de E2C dado que su desarrollo es reprimido por la presencia de ADN específico. Inducida por temperatura, la polimerización de E2C se desencadena mediante la formación de un núcleo estructurado y finaliza en un paisaje morfológico de oligómeros solubles con propiedades amiloideas. El estudio computacional de la interacción E2-ADN de todos los tipos de HPV mucosos indica que las proteínas E2 de distintos tipos poseen propiedades de discriminación de ADN similares. Las diferencias en la interacción E2-ADN se encuentran principalmente en la secuencia de ADN específico, en acuerdo con el análisis de la unión de E2C a diferentes sitios mediante titulaciones isotérmicas de calorimetría. Basado solamente en la secuencia de ADN, logramos predecir diferencias en la energía de unión las cuales se ordenaron en seis grupos de afinidades discretas. Ciertas jerarquías de afinidades como también distancias entre sitios y presencia de sitios de metilación se encontraron estadísticamente asociados con tipos de HPV propensos a ser malignos. Estudiamos la estabilidad y la propiedad de discriminación de secuencia de cinco proteínas E2C homólogas de identidad de secuencia proteica en el rango 44% al 77%. La desnaturalización al equilibrio y la cinética de desplegado en cloruro de guanidinio mostró que todos los dominios son homodímeros estables que se desnaturalizan vía un mecanismo de dos estados. Medimos la unión a distintos sitios de ADN y confirmamos que el mecanismo de unión para todos los complejos es el mismo en base a una genuina compensación entropico-entálpica. Nuestros resultados muestran que la inusual estructura de barril-beta de E2C puede acomodarse a numerosas mutaciones manteniendo las propiedades cruciales de conformación y función. Pero la unión cooperativa a sitios adyacentes de ADN mostró que los componentes entálpicos-entrópicos de la reacción como la deformación del ADN puede divergir entre distintas homólogas en acuerdo con un fuerte acoplamiento entre la dinámica global de la proteína y el reconocimiento del ADN.

Palabras clave: papilomavirus humano, Proteína E2 regulador maestro, Interacciones proteína-ADN, Biofísica.

Resumen en inglés

“E2, transcription and replication regulator of human papillomavirus: protein-DNA interaction, conformational dynamics and evolutionary divergence”

More than 50 Human papillomavirus (HPV) types infect mucosal epithelia causing a variety of benign and malign lesions. The C-terminal domain of HPV E2 protein (E2C) discriminates between four highly similar cognate DNA sites in order to regulate the virus life cycle by activating replication and repressing and/or activating transcription. In order to try to understand the nature of the interaction between E2C and DNA, we performed a detailed analysis of the complex using bioinformatics tools and the information content on DNA sequences and using biophysical techniques and recombinant engineered and homologous E2C proteins and synthesized specific and nonspecific DNA sites. Before studying the role of the E2C-DNA interaction, we focused on the nonfunctional polymerization process of E2C as its development is repressed by the presence of DNA binding sites. Triggered by heat, the polymerization of E2C starts by formation of a stable and structured nucleus and ends in an organized morphology landscape of soluble amyloid-like oligomers. The computational study of the E2-DNA interaction of all mucosal HPV types, indicates that E2 proteins have similar DNA discrimination properties. Differences in E2-DNA interaction among HPV types lie mostly in the target DNA sequence, in agreement with the analysis of binding of E2C to different sites using isothermal titration calorimetry. Based on DNA sequences only, we could predict differences in binding energies which clustered into six discrete affinity hierarchies. Finally, certain distances between sites, affinity hierarchies and their eventual changes upon methylation, are statistically associated with high-risk types. We studied the stability and DNA discrimination properties of five homologous E2C proteins with sequence identities ranging from 45% to 77%. Equilibrium denaturation and unfolding kinetics in guanidinium chloride showed that all five domains are very stable homodimers that unfold via a two-state mechanism. We used isothermal titration calorimetry to follow binding of the five proteins to three different DNA sites. Genuine enthalpy-entropy compensation confirms that the binding mechanism is the same for all complexes. The five domains have nearly identical capacities for DNA sequence discrimination despite the high degree of sequence divergence. Our results show that the unusual E2C beta-barrel can accommodate many mutations while retaining its crucial conformational and functional properties. But cooperative binding to tandem DNA E2 site showed that the enthalpic-entropic components of the reaction and DNA deformation can diverge in agreement with a strong coupling between global dynamics and DNA recognition.

Keywords: Human papillomavirus, E2 master Regulator Protein, Protein-DNA Interactions, Biophysics.

Agradecimientos

A Gonzalo por la confianza y la fuerza que me dio a lo largo de estos años de tesis doctoral en su laboratorio. Por haber sido un ejemplo de empuje y positivismo, por su inagotable creatividad y grandeza.

Al financiamiento del CONICET y de la Universidad de Buenos Aires. A la casa de estudios Fundación Instituto Leloir. A las entidades que costearon cursos y congresos, AMSUD-Pasteur, SBBq y SAB. A la comunidad científica en general que fomenta el acceso gratuito de la información.

A Nacho, mi gurú científico y amigo personal, por compartir lucidez, pasión y criterio, por hacer más divertidos los días en el laboratorio.

A los *pipeteros* (*carajo*) del laboratorio, por estar cerca en el aliento de todos los días y mostrarme las múltiples caras de este oficio. A Marce, Kari, Santi, Clarita, Leo, Nacho, Mari-sol-kena-ángeles, María-Laura-Marta-Gaby, Ángel, Gabi, Fer, Esteban, Cris, Lucía, Sebas, Pauli, Dieg, Juan, Eleo, Dia, Ale, Ricky y a las secretarias Lili y Agus.

A mantenimiento, administrativos, limpieza, droguero, técnicos, becarios, investigadores de la FIL. A todos los que compartieron conmigo sonrisas, miradas, emociones, ideas, rumores y enigmas en los pasillos, mesadas, ascensores, subsuelo, baños, droguero, cursos, veredas, biblioteca, terraza y parque, me incitaron a regresar cotidianamente al trabajo y con ganas.

A los que leyeron, corrigieron, sugirieron muy amablemente este manuscrito, Ángel, Lucía, Nacho, Leo y a mi amor Marian.

A la muchachada de desagradecidos, por hacerme sentir cómodo y con muchos hermanos.

A mi familia, los Dellarole, hermosa Nena y profe Ale y a los Hughes, tíos y primos.

A mis hermanos de sangre Maxi, Kity y Nico, por su respaldo incondicional.

En memoria de Male, por mostrarme de donde vengo.

A mi vida Marian, por definir el sentido y el punto final de esta tesis, un futuro juntos.

Dedicatoria.

A mis viejos, por enseñarme lo fundamental
No temer a las tormentas
Por todo su amor

Índice

Abreviaturas.	13
Producción científica.	14
Primera Sección: Introducción.	16
- <i>Prefacio.</i>	17
- <i>Papilomavirus humano.</i>	19
- <i>Estructura del genoma y descripción del ciclo infeccioso de papilomavirus.</i>	23
- <i>E2 una proteína multifuncional.</i>	25
- <i>El dominio C-terminal de la proteína E2, topología y plegamiento.</i>	28
- <i>Reconocimiento cooperativo de ADN.</i>	31
Segunda Sección: Materiales y Métodos.	35
1) Materiales	
- <i>Drogas y equipamiento lógico.</i>	36
- <i>Proteínas recombinantes.</i>	36
- <i>Oligonucleótidos sintéticos.</i>	39
2) Métodos	
- <i>Estudios de agregación.</i>	40
- <i>Análisis morfológico.</i>	43
- <i>Bioinformática.</i>	44
- <i>Calorimetría y espectroscopía para complejos binarios.</i>	48
- <i>Estudios de estabilidad.</i>	50
- <i>Calorimetría y espectroscopía para complejos ternarios.</i>	52
Tercera Sección: Resultados	55
Capítulo I. “Mecanismo de formación de protofibras amiloides en E2C”	
- <i>Estados de agregación del dominio de unión a ADN E2C.</i>	57
- <i>scE2C como herramienta para el estudio de amiloidosis.</i>	59
- <i>El tratamiento térmico induce en scE2C un estado de tipo amiloideo.</i>	60

- <i>Diámetro hidrodinámico de los agregados amiloides de scE2C.</i>	63
- <i>Caracterización morfológica de los agregados amiloides de scE2C.</i>	66
- <i>Mecanismo de polimerización de scE2C</i>	69
- <i>La nucleación y la elongación están acopladas.</i>	74
- <i>Caracterización morfológica por microscopía electrónica.</i>	76

Capítulo II. “Análisis estadístico de la discriminación de secuencias específicas de ADN por la proteína E2”

- <i>Conservación de E2 en la hélice de unión a ADN.</i>	80
- <i>Conservación y asimetría de los cuatros sitios de unión.</i>	81
- <i>Posible sitio de metilación conservado dentro del sitio de unión de E2.</i>	84
- <i>Predicción computacional de la afinidad de los dominios de unión de E2.</i>	86
- <i>Seis jerarquías de afinidad prototípicas en alfa papilomavirus.</i>	88
- <i>Asociación entre propiedades moleculares y epidemiológicas.</i>	90

Capítulo III. “Origen termodinámico de la discriminación de secuencias específicas de ADN por la proteína E2”

- <i>Dependencia de la concentración de fosfato en la unión E2C-16 y BS2.</i>	94
- <i>Bases termodinámicas de la discriminación de secuencia de ADN.</i>	99
- <i>Lectura indirecta del lado de la proteína, dependencia del pH.</i>	104

Capítulo IV. “Plegamiento, estabilidad conformacional y reconocimiento de ADN de 5 dominios de E2C correspondientes a los tipos de papilomavirus humano de mayor relevancia”

- <i>Contexto evolutivo y estructura primaria de 5 dominios E2C recombinantes.</i>	110
- <i>Análisis espectroscópico de la estructura secundaria, terciaria y cuaternaria.</i>	115
- <i>Desnaturalización al equilibrio, estabilidad conformacional de los dominios.</i>	118
- <i>Cinéticas de desnaturalización.</i>	122
- <i>Interacción entre E2C y BS2, cambios conformacionales en el ADN.</i>	125
- <i>Interacción entre E2C y ADN, termodinámica de unión.</i>	127

Capítulo V. “Bases termodinámicas del reconocimiento cooperativo del origen de replicación y transcripción de papilomavirus”

- <i>Abordaje experimental y modelo de unión.</i>	134
- <i>Calorimetría de titulación isotérmica para complejos ternarios.</i>	138
- <i>Termodinámica del reconocimiento cooperativo.</i>	140
- <i>Cambios conformacionales en la interacción entre E2C y DBS al equilibrio.</i>	146
- <i>Cambios estructurales del DBS en función del tiempo.</i>	149
Cuarta Sección: Discusión.	151
- <i>Plasticidad conformacional de E2C.</i>	152
- <i>Discriminación de secuencias específicas de ADN por la proteína E2.</i>	157
- <i>Estudios termodinámicos, estabilidad y reconocimiento de ADN.</i>	163
- <i>Bases termodinámicas para el reconocimiento cooperativo de ADN.</i>	169
- <i>Epílogo.</i>	175
Quinta Sección: Referencias.	177
Fin.	200

Nota: en el texto, las Figuras que se mencionan con el número romano pertenecen a otro Capítulo.

Abreviaturas.

A, T y C, G	bases nitrogenadas adenina, timina y citosina y guanina
ADN	ácido desoxirribonucleico
AFM	microscopía de fuerza atómica
ANS	8-Anilino-1-naftaleno ácido sulfónico
Bis-Tris	bis(2-Hidroxietil)imino-tris(hidroximetil) metano
BS	sitios específicos de ADN de unión de E2
CD	dicroísmo circular
CM	centro de masa
DBS	sitios BS1 y BS2 dispuestos en tándem
DLS	difusión dinámica de la luz
DLS	dispersión dinámica de la luz
DTT	di-tiotreitol
e.g.	por ejemplo
EDTA	ácido tetra-acético etilendiamina
EMSA	ensayo de cambio en la movilidad electroforética
FI	Fuerza iónica
FITC	tiocianato de fluoresceína
GdmCl	cloruro de guanidinio
<i>genus</i>	del latín, género
HPV	papilomavirus humano
IPTG	isopropyl β -D-1-thiogalactopyranoside
ITC	calorimetría de Titulación Isotérmica
K_D	constante de disociación
LCR	región génica larga de control en papilomavirus
LEM	método de extrapolación lineal
<i>m</i>	cooperatividad del proceso de desnaturalización
MALDI-TOF	desorción/ionización láser asistida por matriz.
OBP	proteína de unión al <i>ori</i>
ORF	marco abierto de lectura
<i>ori</i>	origen de replicación
PAGE	electroforesis en gel de poliacrilamida
<i>pI</i>	punto isoeléctrico
PCR	reacción en cadena de la polimerasa
PDB	base de datos de estructuras de proteínas
RMN	resonancia magnética nuclear
rpm	revoluciones por minuto
SC	cadena simple
SDS	dodecil sulfato de sodio
U.A.	unidades de Absorbancia
ua	unidades arbitrarias
UV	ultra violeta
ΔG	cambio de energía libre de Gibbs
ΔH	cambio de Entalpía
ΔS	cambio de Entropía
$\epsilon_{280 \text{ nm}}$	coeficiente de extinción molar.

Producción científica.

La presente tesis generó los siguientes artículos publicados:

2010: "Thermodynamics of cooperative DNA recognition at a replication origin and transcription regulatory site." **Biochemistry**, **49**:10277-86. Dellarole M, Sánchez IE, de Prat Gay G.

2009: "Indirect DNA readout on the protein side: coupling between histidine protonation, global structural cooperativity, dynamics and DNA binding of the human papillomavirus type 16 E2C domain." **Journal of Molecular Biology**, **388**:327-44. Eliseo T*, Sánchez IE*, Nadra AD, Dellarole M, Paci M, de Prat Gay G, Cicero DO.

2008: "Comprehensive comparison of the interaction of the E2 master regulator with its cognate target DNA sites in 73 human papillomavirus types by sequence statistics." **Nucleic Acids Research**, **36**:756-69. Sánchez IE*, Dellarole M*, Gaston K, de Prat Gay G.

La presente tesis generó los siguientes trabajos para su publicación:

"Long-range structural perturbation in a transcriptional regulator is proportional to DNA sequence discrimination" Con la participación de Nadra AD, Paci M, Eliseo T, Sánchez IE, Dellarole M, Buslje CM, Cicero DO y de Prat Gay G.

"Hierarchical morphology of soluble amyloid-like oligomers induced by heat-denaturation at physiological conditions of an engineered monomeric protein." Dellarole M y de Prat Gay G con la participación de Sánchez IE y Bebeacua C.

"Strict conservation of stability and DNA sequence discrimination in spite of extensive sequence divergence in five homologous viral ori binding domains." Dellarole M y de Prat Gay G, con la participación de Accialini P y Sánchez IE.

Los siguientes artículos publicados se relacionan directamente con la presente tesis y fueron realizados previo y/o durante el desarrollo de la presente la tesis en colaboración con otros proyectos del laboratorio:

2010: "Experimental snapshots of a protein-DNA binding landscape." **Proceedings of the National Academy of Sciences of the USA**, **107**:7751-6. Sánchez IE*, Ferreiro DU*, Dellarole M, de Prat Gay G.

2007: "Increased stability and DNA site discrimination of "single chain" variants of the dimeric beta-barrel DNA binding domain of the human papillomavirus E2 transcriptional regulator." **Biochemistry**, **46**:12441-50. Dellarole M*, Sánchez IE*, Freire E, de Prat-Gay G.

2006: "Structural and thermodynamic basis for the enhanced transcriptional control by the human papillomavirus strain-16 E2 protein." **Biochemistry**, **45**:6551-60. Cicero DO, Nadra AD, Eliseo T, Dellarole M, Paci M, de Prat-Gay G.

2005: "Free energy contributions to direct readout of a DNA sequence." **Journal of Biological Chemistry**, **280**:32480-4. Ferreiro DU, Dellarole M, Nadra AD, de Prat-Gay G.

*Contribución igualitaria.

Primera Sección

Introducción

Information is not knowledge

Knowledge is not wisdom

Wisdom is not truth

Truth is not beauty

Beauty is not love

Love is not music

Music is the best

Frank Zappa

Prefacio.

A principios de la década de 1970,

el investigador Paul Berg desarrolla un método eficiente con el cual logra insertar genes del fago lambda, incluyendo el operón galactosa, al genoma del virus SV40 (Jackson *et al.* 1972). Nace entonces la primera construcción genética artificial dando comienzo a la era del ADN recombinante. La manipulación genética de Paul Berg hubiese sido imposible sin el previo descubrimiento fortuito de las enzimas de restricción por Werner Arber (Arber 1965). Los hallazgos de Berg dieron a su vez un salto clave gracias al trabajo posterior de Stanley Cohen y Herbert Boyer, quienes insertaron ADN de rana en un plásmido bacteriano y lo introdujeron en la bacteria *Escherichia coli* (Morrow *et al.* 1974). Estos organismos se duplican cada 20 minutos y con ellos el ADN foráneo, permitiendo un eficiente clonado del ADN recombinante. En base a estos descubrimientos, Herbert Boyer creó en 1976 la empresa *Genentech*, fundadora de la industria biotecnológica.

La importancia de los avances científicos acumulados en la década del 70 dieron comienzo a la era de la biología molecular, de dogma central diagramado por Francis Crick (Crick 1958; Crick 1970), y gestaron grandes expectativas para un futuro cercano (Crick 1970). A su vez, los flamantes desarrollos generaban importantes desafíos e interrogantes, los cuales fueron plasmados principalmente en la conferencia de Asilomar llevada a cabo en el año 1975, en California (Berg *et al.* 1975). La conferencia, convocada por Paul Berg, se originó para evaluar los potenciales riesgos de la manipulación genética, siendo el principal temor la posibilidad de crear una bacteria virulenta para el hombre al insertarle ADN de virus cancerígeno. En la conferencia se decidió continuar con las investigaciones de manipulación genética dentro de ciertas reglas de seguridad, consolidándose la disciplina que consiste en aislar, copiar y transferir genes de un organismo a otro (Berg *et al.* 1975). En 1977, con el fin del tratamiento de la diabetes, se comienza a producir insulina en forma recombinante (Ullrich *et al.* 1977), convirtiéndose ésta en la primer droga terapéutica genéticamente manipulada (Humilin, *Genentech*, 1978). Frente al potencial uso para el bienestar de la humanidad, la manipulación genética deja de ser controversial (Berg *et al.* 1995). Hoy en día, luego de más treinta años del

comienzo de la disciplina, los beneficios y avances científicos generados a partir del ADN recombinante se asientan en una lista extremadamente difícil de compilar dada su cantidad y variedad, desde métodos de diagnóstico y prevención a tratamientos de enfermedades humanas mediante vacunas, hormonas y diversos agentes terapéuticos (Berg *et al.* 2010).

Una de las múltiples aristas de esta explosión de conocimientos y beneficios traídos por la técnica de ADN recombinante se refleja claramente en el crecimiento exponencial del número de estructuras de proteínas resueltas por año. De 13 estructuras proteicas resueltas en el año 1976 se pasó a 7400 estructuras resueltas en el año 2010, acumulando un total de 65 mil estructuras proteicas resueltas en 34 años y de las cuales más de 1500 corresponden a estructuras de complejos proteína-ADN. Dichas estructuras fueron provistas por difracción de rayos X (87%), resonancia magnética (12%) y microscopía electrónica (1%) (PDB, www.pdb.org). Las tres técnicas de resolución atómica requieren de la producción de proteínas con alto grado de pureza y en abundancia, dos aspectos logrados gracias al desarrollo de la tecnología del ADN recombinante. Gracias a ello, los estudios de función proteica se racionalizan en conjunto con información de estructura tridimensional. Otro aspecto contundente facilitado a partir del uso de enzimas de restricción, de ADN recombinante y de clonado de ADN es el secuenciado del mismo. Hoy en día contamos con el acceso a la secuencia de más de 12 mil genomas completos, más de 78 millones de genes y más de 37 millones de proteínas (NCBI, www.ncbi.nlm.nih.gov). En forma sinérgica con la resolución de estructuras a nivel atómico, el acceso a semejante cantidad de secuencias provenientes de una gran variedad de organismos permite profundizar la comprensión de estudios de función proteica a un nivel evolutivo, extrayendo información a partir de la variabilidad de secuencia.

Las técnicas de biología molecular desarrolladas en los años 70 constituyen los pilares directos e indirectos de la presente tesis doctoral. Utilizaremos bases de datos de secuencias genómicas para extraer información biológica y con ella tratar de comprender la interacción entre una proteína viral y su ADN blanco. Emplearemos estructuras de ADN, de proteínas y de complejos proteína ADN resueltas por cristalografía, resonancia magnética nuclear y microscopía electrónica para profundizar nuestros resultados. Purificaremos en forma recombinante en la bacteria *Escherichia coli*, proteínas de un virus cancerígeno con el fin de comprender cómo la interacción de estas proteínas con el ADN, su dinámica conformacional y su divergencia evolutiva modula el desarrollo del ciclo infeccioso de este patógeno, el

papilomavirus humano.

Papilomavirus humano.

Los papilomavirus, PV, son patógenos que infectan el tejido epitelial y que se encuentran sorprendentemente dispersos en la población humana (Munoz *et al.* 2003; Bosch *et al.* 2006). En tal medida que todas las poblaciones humanas han sido infectadas con papilomavirus, concediendo el uso de las variantes genéticas del virus para estudiar las migraciones y la dinámica poblacional del hombre (Chan *et al.* 1992). Además de encontrarse dispersos en el hombre, los PV infectan una gran proporción del reino animal, extendiendo su evolución en todos los amniotas, es decir mamíferos, aves (Tachezy *et al.* 2002) y reptiles (Herbst *et al.* 2009). Se aislaron 189 variantes genéticas o tipos distintos de PV, los cuales se agrupan, de acuerdo a grado de identidad en secuencia nucleotídica, en 29 géneros o *genera* definidos (Bernard *et al.* 2010). La gran mayoría de tipos fueron aislados en tejidos humanos, los papilomavirus humanos (HPV), con 120 tipos distintos que representan el 70% del conjunto total de genomas aislados del virus. De los 29 *genera*, los HPVs se encuentran en solo cinco mientras que la totalidad de los mamíferos se agrupan en 20 *genera* (Bernard *et al.* 2010). En la **Figura 1**, tomamos de (Bernard *et al.* 2010) el árbol de filiación genética de todos los PV secuenciados a la fecha de finalización de escritura de la presente tesis. Podemos apreciar que de los cinco *genera* que presentan tipos de HPV, la mayoría se encuentra repartidos en tres, el *genus* alfa y el *genus* beta y el *genus* gamma (**Figura 1**). En la presente tesis nos focalizaremos al estudio PV del *genus* con mayor cantidad de tipos y de mayor relevancia medico-sanitaria para el hombre, el *genus* alfa. Al final de la tesis discutiremos la relevancia del *genus* beta, el segundo en número de tipos aislados y de creciente relevancia epidemiológica.

Todos los tipos de HPV aislados de epitelio mucoso pertenecen al *genus* alfa junto a doce tipos de HPV aislados de epitelio cutáneo y dos papilomavirus de simios (de Villiers *et al.* 2004; Bravo *et al.* 2007; Bernard *et al.* 2010). Los tipos de HPV mucosos son los agentes etiológicos de cáncer vaginal, anal, peneano y oral-faringeal (Bosch *et al.* 2006). Los tipos de HPV difieren ampliamente en su potencial oncogénico de acuerdo a evidencias epidemiológicas (Munoz *et al.* 2003) y se separan en dos grupos principales. 19 tipos pertenecen al grupo clasificados de “alto-riesgo” para el desarrollo de cáncer: los tipos HPV16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68a, 73, 82 y 82subtipo, 13 tipos pertenecen al grupo clasificado de “bajo-riesgo”

para el desarrollo de cáncer: los tipos HPV6, 6a, 6b, 11, 40, 42, 43, 44, 54, 61, 70, 72 y 81 (Figura 1).

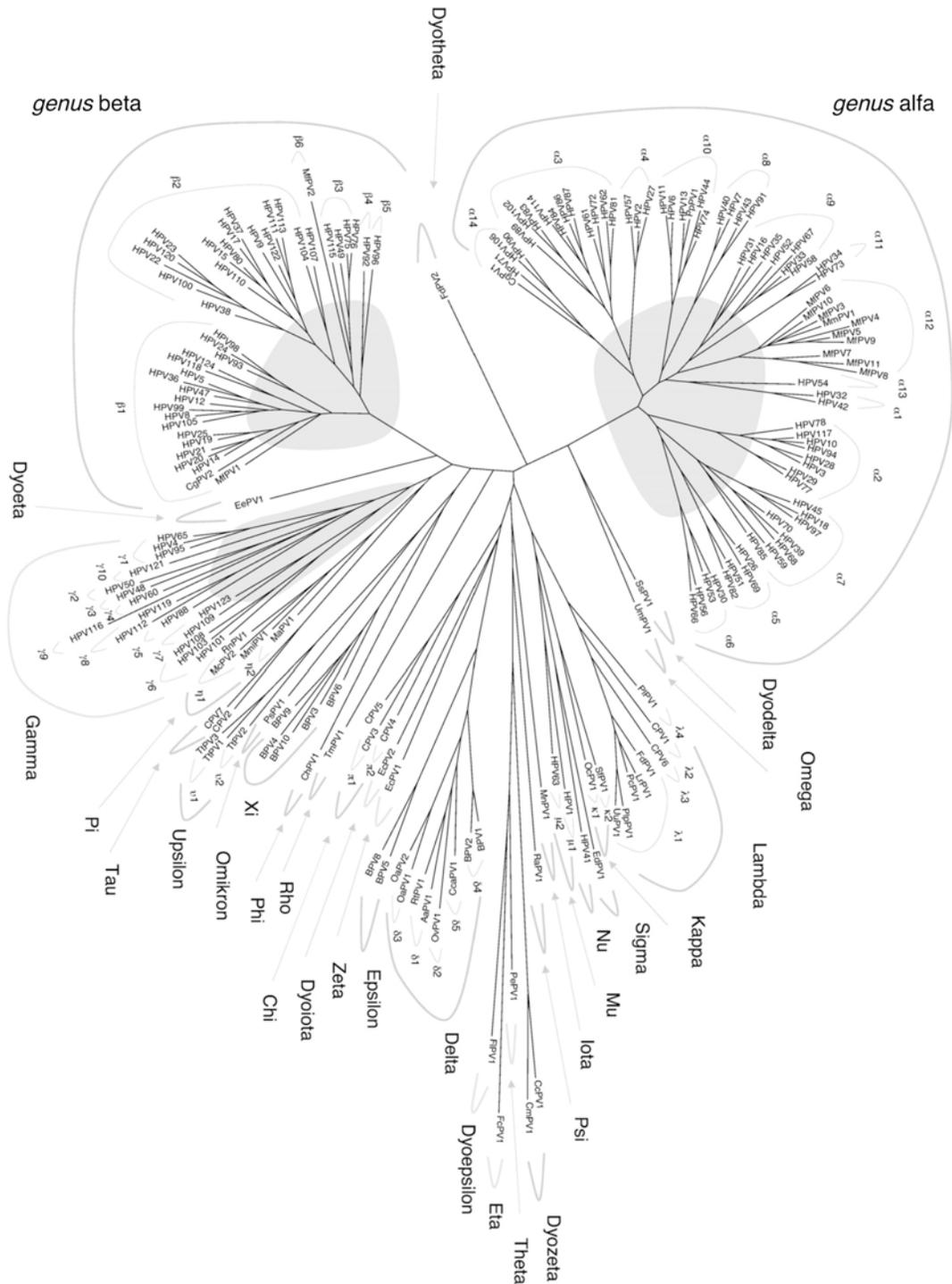


Figura 1. Árbol filogenético de papilomavirus. Se representan 29 *genera* que abarcan 189 tipos virales de los cuales el 70% fueron aislados en humanos (HPV), el 29% restante fueron aislados en otros mamíferos y el resto minoritario en reptiles y pájaros. En gris se destacan los *genera* con mayor presencia de tipos humanos, el *genus* alfa, el *genus* beta y el *genus* gamma, también presentan tipos de HPVs el *genus* nu y el *genus* mu. Adaptado sin permisos de (Bernard *et al.* 2010).

El virión de PV posee una única copia de ADN doble cadena circular de ~8000 pares de bases contenidos en una cápside icosaédrica proteica sin envoltura (**Figura 2, arriba**). La cápside, de ~50 nm de diámetro, presenta 72 capsómeros compuestos por dos proteínas estructurales: L1 y L2 (Chen *et al.* 2000; Modis *et al.* 2002) (**Figura 2, arriba**). La transmisión del virus en humanos ocurre por contacto directo y contacto sexual (**Figura 2, abajo**) y durante el nacimiento mediante contacto con el canal de parto. La actividad sexual es el mayor factor de riesgo de contagio (WHO 2007). La infección mediada por el virus produce una variedad de enfermedades, que pueden derivar de verrugas benignas a cáncer. En el mundo, el cáncer de cuello de útero, con ~medio millón de nuevos casos por año y ~un cuarto de millón de muertes por año (WHO 2007), es la segunda causa de muerte por cáncer en mujeres, y el agente etiológico es exclusivamente, en un 99% de los casos, la infección con tipos de HPV del *genus* alfa de la categoría de 'alto riesgo' (zur Hausen 1996). En algunos países pobres, el cáncer del cuello de útero lidera como causa de muerte por cáncer, por ejemplo en la India con ~70000 muertes por año (IARC 2007). La cápside de PV (**Figura 2, arriba**), es el principal exponente del virus durante la infección y es por ello que las vacunas profilácticas utilizan como componente inmunogénico a L1, la principal proteína expuesta en la superficie del virión.

Las vacunas profilácticas, Gardasil[®] (Merck) contra los tipos HPV6, 11, 16 y 18 y Cervarix[®] (GlaxoSmithKline) contra los tipos HPV16 y 18 fueron desarrolladas a principios de los años 1990 (McNeil 2006) y se encuentran desde 2007 en el mercado a un costo de >100 US\$ la dosis. Las vacunas protegen contra el ~70% de cáncer cervical. Paradójicamente en los países más pobres (con salarios de 1 a 4US\$/día) dónde acontecen ~80% de las muertes por cáncer de cuello de útero del mundo, el tratamiento resulta económicamente inaccesible (Cohen 2005). En Argentina, ~5000 mujeres son diagnosticadas cada año con cáncer cervical y ~2000 mueren cada año de este flagelo (WHO 2007). Al momento de finalización de la escritura del presente manuscrito, el tratamiento contra HPV ha sido declarado dentro del plan de vacunación nacional, siendo gratuita la vacunación para mujeres de 11 años de edad. Esta medida pionera en Latinoamérica favorece principalmente a las regiones más pobres y susceptibles a la enfermedad, como es el Noroeste argentino. Sin embargo, las vacunas existentes no protegen contra infecciones de otros tipos de alto o bajo riesgo, no curan a las millones de personas ya infectadas, para las cuales no hay tratamiento, y el número de variantes genotípicas crece año a año (Bernard *et al.* 2010). Por esta razón, aún en pacientes inmunizados, los virólogos expertos del campo

recomiendan realizar chequeos frecuentes mediante la prueba de Papanicolaou (Papanicolaou 1942). Por lo tanto, a costas de la existencia de vacunas profilácticas, impera aún la necesidad de comprender el modo de acción y la oncogenicidad de los papilomavirus con el fin de desarrollar tanto estrategias profilácticas más eficientes y económicas como drogas antivirales.

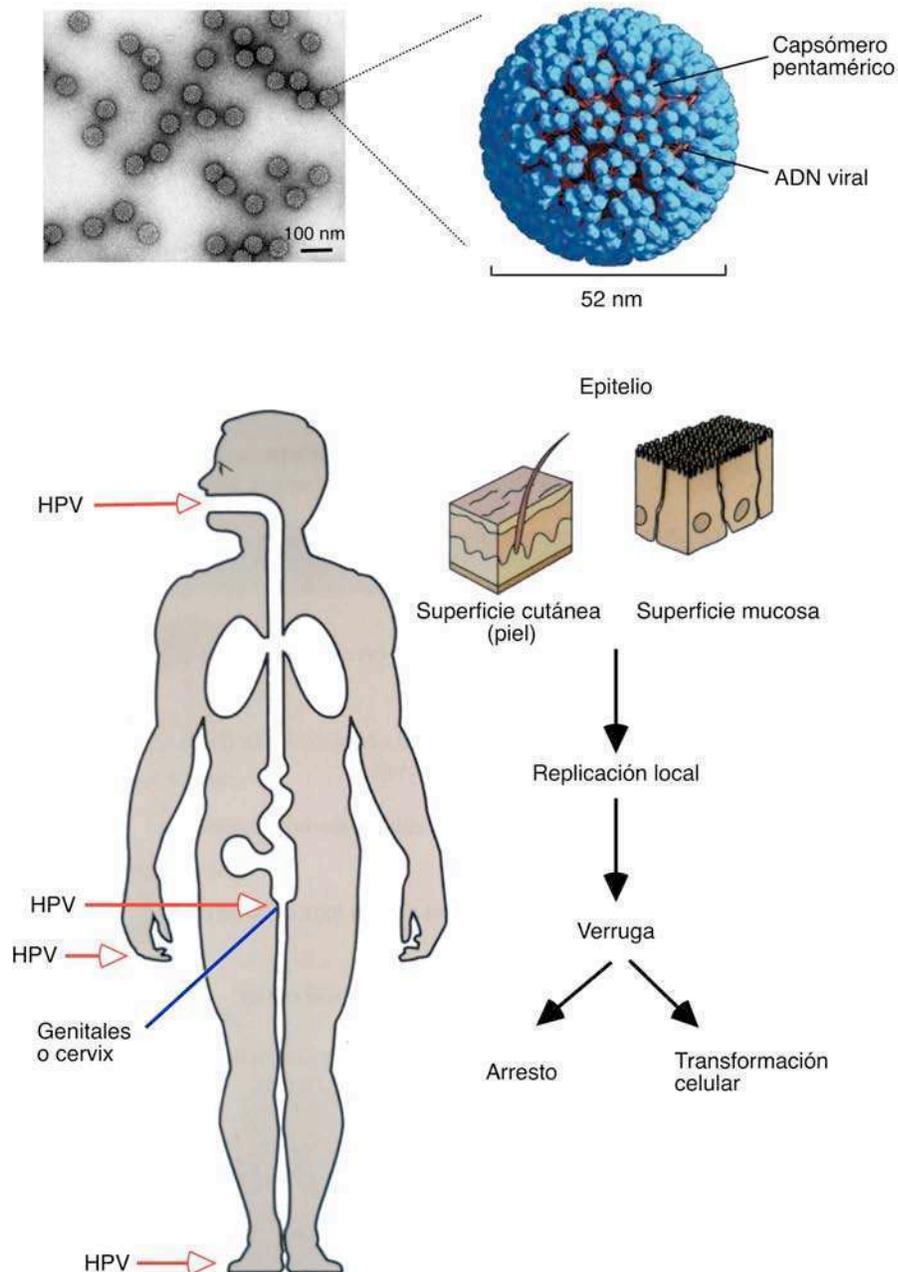


Figura 2. Estructura del virión y epidemiología de HPV. Arriba: partículas de HPV provenientes de muestras de verrugas cutáneas reveladas por microscopía electrónica (imagen obtenida modificada sin autorización de (Orth 1999)). (Chen *et al.* 2000; Modis *et al.* 2002) En azul: los capsómeros hexavalentes, constituidos por las proteínas virales L1 y L2. Abajo: esquema del mecanismo de infección papilomavirus en humanos. Imagen obtenida y modificada sin permisos de (Flint *et al.* 2008).

Estructura del genoma y descripción del ciclo infeccioso de papilomavirus.

El pequeño genoma de papilomavirus, esquematizado en la **Figura 3, derecha**, presenta cuatro genes extensos, E1, E2, L1 y L2, cuatro marcos de lectura abiertos pequeños, E4, E5, E6 y E7 y una región no codificante, LCR, donde se encuentran múltiples sitios de unión de proteínas involucradas en la regulación de la replicación y expresión de las proteínas virales. El nombre de las proteínas de PV se corresponde con la fase temporal del ciclo infeccioso del virus: E por *Early* o proteínas expresión temprana y L por *Late* o proteínas de síntesis tardía. El ciclo infeccioso del virus, es decir la finalización de un ciclo de replicación desde la entrada a la síntesis de nuevas partículas, presenta un patrón de expresión de proteínas de virales que depende del estadio de diferenciación epitelial (**Figura 3, izquierda**).

El desarrollo del ciclo infeccioso del PV puede ser descrito a grandes rasgos de acuerdo a la función de las proteínas del virus. En una primera instancia del ciclo se expresan mayoritariamente las proteínas E6 y E7 (**Figura 3**). Ambas son consideradas proteínas promiscuas, ya que presentan la capacidad de interactuar con múltiples proteínas del huésped. Esta particularidad es conferida por aspectos estructurales particulares, estudiados ampliamente en nuestro laboratorio (Alonso *et al.* 2006; Garcia-Alai *et al.* 2007; Heer *et al.* 2011). El principal blanco de E6 y de E7 consiste en proteínas involucradas directamente en el ciclo celular del huésped (Howley 2001). En células infectadas con HPV, los niveles de p53 se mantienen bajos gracias a la interacción con E6, la cual junto con la ubiquitina ligasa celular E6-AP degradan a p53 (Howley 2001). Se ha reportado también que E6 interactúa con como ser paxilina, hDLG, IRF-3, Bak, CBP/p300 (Howley 2001). A su vez, la proteína E7 interactúa con varias proteínas involucradas en la regulación del ciclo celular, entre otros el supresor de tumores retinoblastoma, p107, p130, el factor de transcripción celular E2F, la ciclinaA (Howley 2001). Esta delicada red de interacción le confiere a E6 junto con E7 la propiedad de ser oncogénica: la expresión desmesurada de ellas las convierte en agentes responsables del proceso carcinogénico de PV. E7 es capaz de immortalizar líneas celulares en cultivo, tanto sola como en cooperación con E6 y/o el oncogen Ras. Se ha reportado también que la expresión de E5 es capaz de transformar células de ratón en cultivo, estimulando el crecimiento celular. Si bien el efecto oncogénico de E5 actualmente se encuentra en discusión, se acepta que coopera con E6 y E7 en la estimulación de la proliferación celular (Howley 2001).

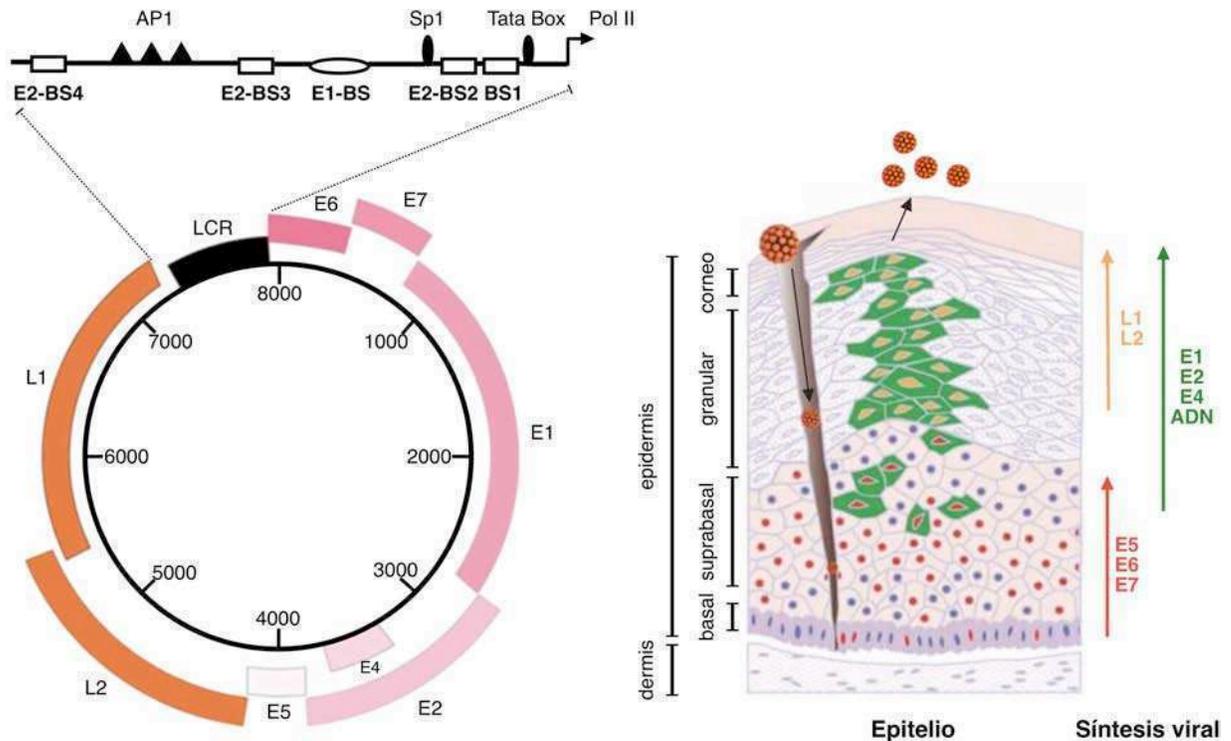


Figura 3. Genoma prototípico y ciclo infeccioso de papilomavirus humano. Derecha: genoma prototipo de HPV de 8000 pares de base de ADN circular. Se indica la posición y extensión de los marcos de lectura abierto de las principales proteínas que codifica: las proteínas tempranas (coloradas en tonos de rosa) E1, E2, E4, E5, E6, E7 y las proteínas tardías L1 y L2 (coloreadas en naranja). Se detalla arriba la región larga reguladora (LCR) indicando la posición de los cuatro sitios de unión de la proteína viral E2: E2-BS, la posición del sitio de unión de la proteína viral E1: E1-BS, algunos de los sitios de unión de factores de transcripción celulares SP1, TATA box, y AP1 y el promotor de expresión de proteínas tempranas Pol II. Izquierda: esquema del ciclo infeccioso de HPV en el epitelio. Se muestra el patrón de expresión de las proteínas virales y de síntesis de ADN viral en los diferentes estratos de diferenciación epidérmica. En rojo la expresión de las proteínas E6 y E7, en verde la expresión de las proteínas E1 y E2 y la replicación del genoma viral y en naranja la expresión de las proteínas tardías L1 y L2, unidades formadoras de la cápside viral. Imágenes adaptadas sin permiso de (Zheng *et al.* 2006) y de (Doorbar 2006).

En segunda instancia se sintetizan las proteínas virales E1 y E2 (**Figura 3**), se reprime la expresión de E6 y E7 y se inicia la replicación del material genético de PV. La proteína E1 es la única proteína codificada en el genoma viral que presenta actividad enzimática de ATPasa y helicasa ATP dependiente, siendo un factor esencial para la replicación del genoma viral (Wilson *et al.* 2002). E1 presenta unión específica, pero débil, a ADN en el origen de replicación del virus (**Figura 3, izquierda, E1-BS**), donde es requerida para la iniciación y la elongación de la síntesis de ADN viral. La unión de E1 a su sitio blanco es facilitada por la proteína E2 (Mohr *et al.* 1990), la cual une fuertemente a los cuatro sitios de ADN (**Figura 3, izquierda, E2-BS**). La replicación del genoma viral *in vivo* requiere de la proteína E2 (Howley 2001). Además de facilitar la síntesis de ADN viral, la proteína E2 es la única proteína viral responsable de la

regulación de la expresión los genes virales, en particular de la represión de los oncogenes E6 y E7. Es por ello que E2 de HPVs de *alto riesgo* parece actuar como supresor de tumores. La disrupción del gen de E2 o la pérdida de su expresión deviene en células que son más propensas a proliferar y, como consecuencia, más propensas a la generación de lesiones tumorales.

El ciclo infeccioso finaliza con el ensamblado del virión y la síntesis de las proteínas que lo componen, L1 y L2 (**Figuras 2 y 3**) que encierran al ADN nuevo, conformando una partícula de PV capaz de infectar nuevos huéspedes. Aparentemente, el genoma de papilomavirus es de origen modular. La divergencia de los 8 marcos de lectura abiertos mencionados presentan distinta historia evolutiva (García-Vallve *et al.* 2005). El protopapilomavirus o virus primitivo estaría compuesto únicamente por los cuatro marcos de lectura extensos y de menor divergencia: E1, E2, L1 y L2 (**Figura 3**), mientras que los genes E5, E6 y E7 habrían sido adquiridos posteriormente en el genoma de PV (García-Vallve *et al.* 2005). El tipo de PV aviar carece de los genes E5, E6 y E7 (Tachezy *et al.* 2002). Estos hallazgos en conjunto con estudios de secuencia, nos ubican en la edad de aparición de los PV en la tierra: es probable que la aparición de los genes de E6 y E7 acompañen a los mamíferos desde su nacimiento, hace ~150 millones de años, mientras que el gen E5 habría sido adquirido más recientemente, hace ~65 millones de años (García-Vallve *et al.* 2005). Así como el estudio de la divergencia de las secuencias genómicas de HPV nos pueden iluminar sobre la historia del virus, el estudio de las secuencias de ADN puede ser clave para entender el funcionamiento del virus. Hasta el momento no se ha estudiado la divergencia en los sitios de unión de ADN de E2, y dada la relevancia de estas secuencias en el ciclo infeccioso del virus, en la presente tesis abordaremos su estudio.

El evento, de apagado de la expresión de las proteínas tempranas E6 y E7 y concomitante encendido de la expresión de las proteínas tardías, ubica al evento de duplicación del material genético en un ‘interruptor’, el cual es accionado por la proteína viral E2, objeto de estudio de la presente tesis. Este tipo de cambio en la expresión génica (*genetic switch* (Ptashne 2004)), es usual que ocurra en general en la replicación de virus de ADN doble cadena (Flint *et al.* 2008).

E2 una proteína multifuncional.

La proteína E2 se une a cuatro sitios conservados (BS), ubicados en la región reguladora del genoma de alfa-papilomavirus (Kalantari *et al.* 2006), numerados de acuerdo a su distancia con el promotor temprano (**Figura 3, izquierda**). La función de cada sitio en los tipos alfa HPV

6, 11, 16, 18 y 31 ha sido estudiada por mutaciones en el sitio y por frenado de la replicación y ensayos de transcripción genética. La unión de E2 a BS1, BS2 y BS3 recluta la proteína E1 al origen de replicación, siendo la unión al sitio BS2 la más importante para esta función, seguido por la unión a BS1 y BS3 (Chiang *et al.* 1992; Remm *et al.* 1992; Lu *et al.* 1993; Demeret *et al.* 1995; Sverdrup *et al.* 1995; Stubenrauch *et al.* 1998; Newhouse *et al.* 2001). La unión de E2 al sitio BS4 induce la transcripción del promotor de proteínas tempranas (Dong *et al.* 1994; Rapp *et al.* 1997; Steger *et al.* 1997), incluyéndose E2 misma y las proteínas oncogénicas E6 y E7. La transcripción desregulada del promotor temprano conlleva a la acumulación de excesivas cantidades de E6 y E7 y está asociada con cáncer. La unión de E2 a BS1, BS2 y BS3 reprime la transcripción del promotor temprano vía el desplazamiento de Sp1 y TBP de sus sitios de unión (**Figura 3, izquierda**), manteniendo los niveles de E6 y E7 bajo control. El sitio de unión BS1 es el más importante para esta unión seguido de BS2 y BS3 (Thierry *et al.* 1991; Dong *et al.* 1994; Demeret *et al.* 1997; Rapp *et al.* 1997; Steger y Corbach 1997; Stubenrauch *et al.* 1998; Newhouse y Silverstein 2001; Soeda *et al.* 2006).

Varios grupos de investigación estudiaron la unión al ADN del dominio E2C de tipos de alfa-HPV 6, 11, 16, 18, 33 y 51 (Sanders *et al.* 1994; Alexander *et al.* 1996; Thain *et al.* 1996; Muller *et al.* 1997; Thain *et al.* 1997; Hines *et al.* 1998; Ferreira *et al.* 2000; Kim *et al.* 2000; Newhouse y Silverstein 2001; Hegde 2002; Dell *et al.* 2003; Ferreira *et al.* 2003; Nadra *et al.* 2004; Zhang *et al.* 2004; Ferreira *et al.* 2005; Blakaj *et al.* 2006; Cicero *et al.* 2006; Hooley *et al.* 2006; Falconi *et al.* 2007). Todos estos dominios se unen a una secuencia diana pseudopalindrómica con la secuencia consenso aACCg(A/T)₄cGGTt, donde las letras en mayúscula indican las bases fuertemente requeridas, las letras en minúscula las bases débilmente requeridas y (A/T)₄ un espaciador de cuatro bases compuesto por A o T (Hegde 2002). Las bases en el espaciador no realizan contactos directos con la proteína pero contribuyen a la energía libre de unión, un fenómeno llamado como “lectura indirecta” (Bedrosian *et al.* 1990; Hegde *et al.* 1998; Hines *et al.* 1998; Rozenberg *et al.* 1998; Kim *et al.* 2000; Hizver *et al.* 2001; Hegde 2002; Dell *et al.* 2003; Zimmerman *et al.* 2003; Byun *et al.* 2004; Zhang *et al.* 2004; Rohs *et al.* 2005; Blakaj *et al.* 2006) (**Figura 4**). La interacción E2-ADN es un importante sistema modelo para el estudio de este tipo de efectos de lectura indirecta en complejos proteína-ADN (Bedrosian y Bastia 1990; Hegde y Androphy 1998; Hines *et al.* 1998; Rozenberg *et al.* 1998; Kim *et al.* 2000; Hizver *et al.* 2001; Hegde 2002; Dell *et al.* 2003; Zimmerman y Maher 2003; Byun y Beveridge

2004; Zhang *et al.* 2004; Rohs *et al.* 2005; Blakaj *et al.* 2006) (**Figura 4**).

Por otro lado, la interacción E2-ADN puede ser regulada por la metilación de los dinucleótidos CG ubicados en dentro del BS. Esta modificación covalente es conocida por reducir la afinidad de la unión de E2 a sus sitios de unión diana (Thain *et al.* 1996) y la actividad transcripcional de la proteína (Kim *et al.* 2003). La metilación *in vivo* puede también modificar la accesibilidad del genoma de HPV mediante remodelación de la cromatina (Rosl *et al.* 1993). Los sitios de unión de E2 de los tipos de HPV 16 y 18 son alcanzados por la maquinaria de metilación en forma regulada de acuerdo al estado de diferenciación de la célula, la integración del genoma viral y la progresión del cáncer (Badal *et al.* 2003; Kim *et al.* 2003; Badal *et al.* 2004; Kalantari *et al.* 2004; Wiley *et al.* 2005; Bhattacharjee *et al.* 2006). No se sabe si la metilación representa un mecanismo de defensa del huésped, una parte integral del ciclo de vida del virus o un evento relacionado con la enfermedad (Badal *et al.* 2003; Kim *et al.* 2003; Badal *et al.* 2004; Kalantari *et al.* 2004; Wiley *et al.* 2005; Bhattacharjee y Sengupta 2006).



Figura 4. Modelo molecular de la curvatura del sitio de unión de E2 para distintos espaciadores. El sitio de unión de E2, posicionado en el medio de 52 pares de base de ADN intrínsecamente recto (en verde), está coloreado de amarillo la secuencia de los hemisitios y de rojo la secuencia del espaciador, para dos espaciadores: el del sitio natural BS4 y del sitio artificial BSR. Este fenómeno de precurvatura del ADN específico influye en la energía de unión de E2 y es llamado “lectura indirecta”. Imagen adaptada de (Zimmerman y Maher 2003).

La unión jerárquica de E2 a sus cuatro sitios diana es considerada crucial para la regulación de la transcripción y replicación del virus (Alexander y Phelps 1996; Dell *et al.* 2003;

Kalantari y Bernard 2006). Dado que los factores celulares que compiten con E2 en la unión al ADN viral poseen una afinidad relativa muy baja (Tan *et al.* 1994), la jerarquía de unión puede en principio ser descripta utilizando sólo las afinidades de E2 para los BS de unión. La jerarquía de unión no se encuentra conservada entre tipos de alto-riesgo y bajo-riesgo, por lo tanto podría estar relacionada con el desarrollo de cáncer (Alexander y Phelps 1996; Dell *et al.* 2003).

Muchos estudios han comparado proteínas de tipos virales prototípicos de alto-riesgo y bajo-riesgo con el objetivo de determinar qué propiedades bioquímicas están asociadas con la progresión de la enfermedad. En el caso de E2, estas propiedades incluyen señales de localización nuclear (Klucevsek *et al.* 2007), mecanismo de unión a ADN (Ferreiro y Prat Gay 2003; Hooley *et al.* 2006), jerarquía de unión a los sitios (Alexander y Phelps 1996; Dell *et al.* 2003) y mayor afinidad por el ADN (Hou *et al.* 2002) y p53 (Parish *et al.* 2006). Todo el trabajo en la interacción de E2 con su ADN diana ha sido enfocado para un acotado número de tipos de HPV. Mientras que es elemental elucidar cómo los tipos más prevalentes y oncogénicos funcionan, la comprensión inteligible de los papilomavirus y su rol en la enfermedad debería incluir todos los tipos conocidos e integrar un rango completo de datos epidemiológicos y bioquímicos (García-Vallve *et al.* 2006).

El dominio C-terminal de la proteína E2, topología y plegamiento.

La proteína E2 es un polipéptido multifuncional que juega un rol crucial en la replicación de HPV (Hegde 2002), la regulación de la transcripción (Hegde 2002; Kalantari y Bernard 2006) y la segregación genómica (Hegde 2002). Es una proteína multi-dominio formada por dos dominios globulares unidos por una región flexible (Hegde 2002). El dominio C-terminal de E2 (E2C) de 80 aminoácidos, funciona como dominio de dimerización y de unión a ADN (McBride *et al.* 1991; Hegde 2002), (**Figura 5**). La proteína E2C une al ADN como homodímero, con una hélice de cada monómero (hélice $\alpha 1$) contactando dos surcos mayores consecutivos de su sitio diana (Kim *et al.* 2000; Hegde 2002), (**Figura 5, izquierda**). Las cadenas laterales contribuyen a la unión en forma aditiva a la energía libre de unión (Ferreiro *et al.* 2005). La posibilidad de contar con diversas estructuras de alta resolución de E2C (Hegde 2002; Nadra *et al.* 2004) permite relacionar parámetros fisicoquímicos de unión a ADN, y apuntar una visión integrada del mecanismo molecular de reconocimiento. En este sentido, la proteína E2 constituye un modelo general de reconocimiento proteína-ADN.

El dominio E2C forma un homodímero que presenta una topología de plegamiento novedosa: el barril β -dimérico (Hegde *et al.* 1992) (**Figura 5**). Esta topología se encuentra conservada en los dominios de unión a ADN de distintos tipos de PV (Hegde 2002) y consiste en un barril formado por ocho láminas beta de las cuales cada monómero aporta cuatro y dos hélices alfa ubicadas hacia el lado externo del barril beta. La topología secundaria del barril β -dimérico se encuentra descrita en la **Figura 5, derecha**. Esta topología conformando un dímero ha sido descrita únicamente para E2 y para dominio de unión a ADN del antígeno nuclear EBNA1 del virus Epstein-Barr (Bochkarev *et al.* 1996). Al igual que E2, EBNA es también una proteína iniciadora de la replicación (OBP, de sus siglas en inglés). Sin embargo E2 y EBNA no poseen homología de secuencia ni comparten especificidad por los sitios de unión a ADN.

Hoy en día contamos con estructuras de alta resolución para los dominios E2C de varios papilomavirus. Cristalográficamente se han resuelto los dominios de los tipos virales HPV6 (Dell *et al.* 2003), HPV16 (Hegde y Androphy 1998; Nadra *et al.* 2004), HPV18 (Kim *et al.* 2000) y HPV31 (Liang *et al.* 1996; Bussiere *et al.* 1998). En co-cristales con ADN los dominios homólogos de HPV6 (Hooley *et al.* 2006) y HPV18 (Kim *et al.* 2000). En todos los casos se conserva la topología de barril β -dimérico, siendo la desviación media de los carbonos alfa (RMSD) de todas las estructuras de E2C de entre 1 y 2 Å. Es decir que la estructura tridimensional de E2C se mantiene casi intacta entre los distintos tipos. Inclusive comparando con estructuras de E2C de tipos virales evolutivamente lejanos como las de papilomavirus bovino BPV1 (Hegde *et al.* 1992; Hegde *et al.* 1998; Veeraraghavan *et al.* 1999).

E2C es además modelo de estudios básicos de plegamiento proteico, en particular de plegamiento de dímeros. Los estudios iniciales mostraron que la dimerización de E2C está acoplada al plegamiento de los monómeros, es decir, en solución sólo pueden encontrarse monómeros desplegados o dímeros plegados (Mok 1996). Los monómeros constituyen dos cadenas polipeptídicas que conforman un sólo dominio de plegamiento independiente. Se describió también que este dímero es muy estable, con una constante de disociación en el orden subnanomolar, estabilizándose aún más a pH neutro. La descripción del camino cinético de plegamiento de E2C identificó la formación de un intermediario monomérico, de rápida formación y desaparición (Mok *et al.* 1996). Se identificó también que este intermediario presenta características que lo identifican como no-nativo (Mok *et al.* 1996; Prat Gay *et al.* 2005). El efecto de la fuerza iónica reveló que el dímero E2C se estabiliza con el aumento de

ésta, y que dicha estabilización resulta mayor aún si se utilizan sales de fosfato. Tanto la unión específica de ADN como la unión inespecífica de heparina estabilizan fuertemente el dímero frente al evento de desplegamiento/disociación inducido por urea (Lima *et al.* 1997).

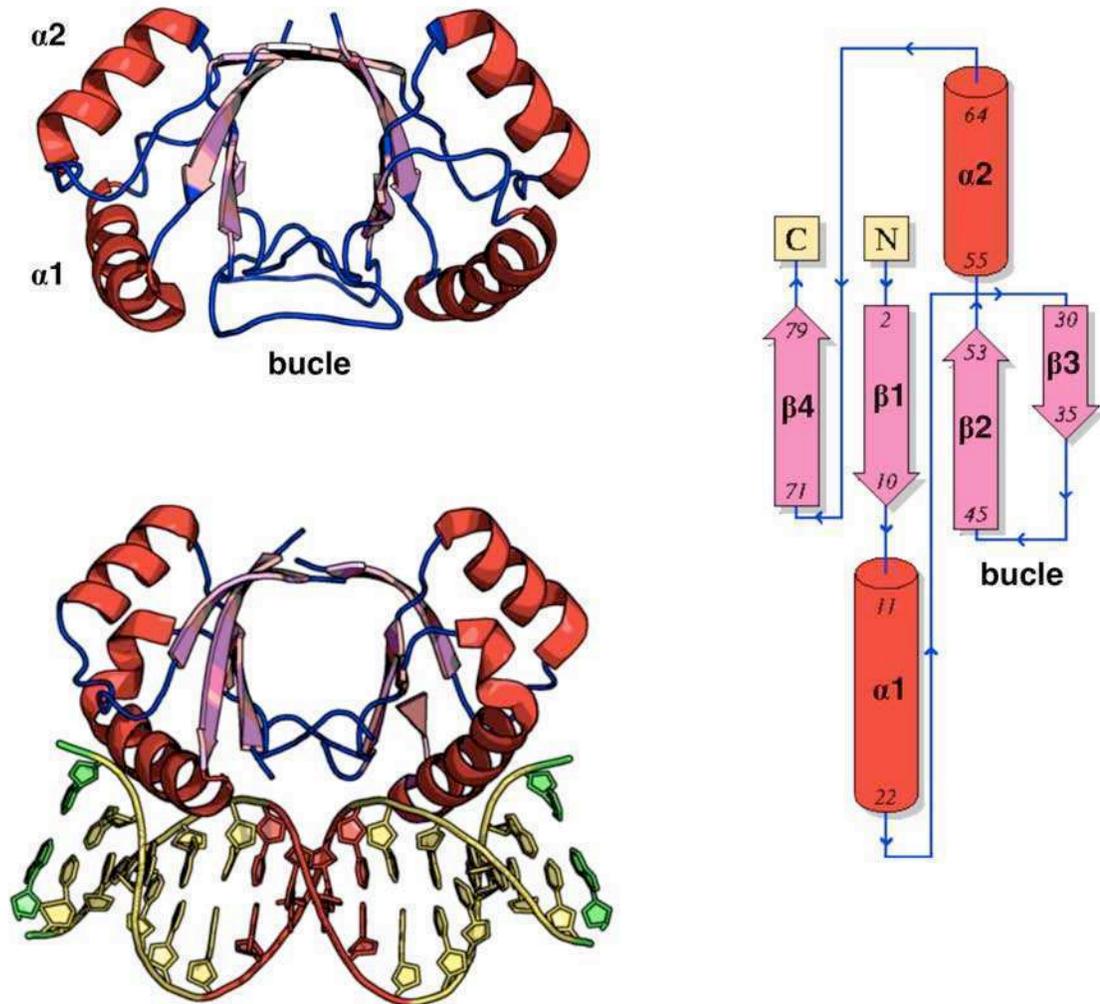


Figura 5. Estructura tridimensional y topología del dominio C-terminal de E2 libre y unido a ADN. **Izquierda:** representación tridimensional de E2C libre (arriba, PDB: 1r8p de E2C-16 resultado por RMN, (Nadra *et al.* 2004)) y unido a ADN específico (abajo, PDB: 1jj4, de E2C-18 resuelto por cristalografía (Kim *et al.* 2000)), ilustrando la estructura secundaria. Imágenes generadas en Pymol (Delano). **Derecha:** Topología del dominio C-terminal de unión al ADN. (PDB: 1zzf, de E2C-16, generada en (Laskowski 2009)). Coloreadas en rojo las hélices alfa, en magenta las laminas beta y en azul los bucles conectores, y para el sitio de ADN de unión de E2C, en amarillo se coloreó la secuencia de los hemisitios (AACCG y CGGTT) y en rojo la secuencia del espaciador (AATT).

E2C también se utilizó como modelo de desnaturalización sin el uso de agentes químicos, como es el caso de la alta presión hidrostática (Foguel *et al.* 1998). Se identificó que los monómeros desplegados por presión presentan estructura residual, y que éste estado es más

compacto que el estado desplegado obtenido por desnaturalización con urea. Los estudios estructurales por RMN sobre el estado desplegado en urea muestran que éste también presenta estructura residual. Fue posible asignar que este estado es monomérico y que las zonas de la hélice $\alpha 1$ y la lamina $\beta 2$ presentan estructuras no-nativas, tipo “giro” (Mok *et al.* 2000). Posteriormente se propuso que una vez unido a ADN, el dímero de E2C puede ser desplegado por presión hidrostática en monómeros que se mantienen unidos a ADN (Lima *et al.* 2000). Por otro lado, estudios de resonancia magnética nuclear permitieron asignar un alto grado de movilidad a las regiones de E2C involucradas en la unión a ADN. Los experimentos de intercambios protón/deuterio realizados sobre el homólogo de HPV31 muestran que tanto el bucle (**Figura 5**), como las hélices de reconocimiento de ADN presentan altos grados de movilidad local cuando el dímero no se encuentra unido a ADN (Liang *et al.* 1996). Experimentos recientes de dinámica de relajación de protones amídicos realizados con el homólogo de HPV16 por nuestro grupo presentan alta flexibilidad local en las zonas de unión a ADN, sugiriéndose además que el efecto podría estar controlado a nivel del extremo N-terminal de la hélice $\alpha 1$ (Cicero *et al.* 2006).

Reconocimiento cooperativo de ADN.

La unión cooperativa de múltiples proteínas a sitios de ADN adyacentes permite regulaciones precisas y económicas de cambios genéticos y del reclutamiento de la maquinaria transcripcional y de replicación (Ptashne 2004). Tres escenarios son posibles: cooperatividad positiva, es decir, la unión de la primera proteína promueve la unión de la segunda; unión aditiva, es decir la unión de la primera proteína es invisible a la unión de la segunda; y cooperatividad negativa, es decir la unión de la primera proteína es adversa a la unión de la segunda (Wyman *et al.* 1990). La termodinámica y los cambios estructurales que acompañan la unión cooperativa de proteínas a ADN son claves a la hora de comprender el mecanismo molecular de reconocimiento y la torsión y apertura del ADN mediado por proteínas. A pesar de su importancia, dichos mecanismos moleculares aún se hallan ampliamente inexplorados.

La replicación semiconservativa del ADN en eucariotas empieza con el reconocimiento de secuencias definidas llamadas origen (*ori*) por proteínas de unión a *ori* (OBPs, acrónimo de sus siglas en inglés). Éstas secuencias definidas son a menudo repeticiones en tándem de un sitio específico. Una vez unidas, las OBPs reclutan helicasas hexaméricas y otras proteínas que

conforman la maquinaria de síntesis de oligonucleótidos activando así la replicación del ADN. La iniciación de la replicación usualmente reprime la transcripción (Prado *et al.* 2005) de forma tal de evitar la colisión de dos maquinarias (Liu *et al.* 1995).

Los virus son las entidades auto-replicas más pequeñas, deben replicar sus genomas para generar nueva progenie. Este evento siempre requiere de por lo menos una proteína viral, en algunos casos muchas proteínas, pero no hay ningún virus que puede hacer todo por sí mismo. Dado que la transcripción y replicación viral se desarrollan usando maquinaria del huésped, el estudio de genética molecular viral dio muchos avances para la comprensión de la transcripción y replicación de genomas eucariotas (Flint *et al.* 2008; Flint *et al.* 2008). Una cuestión muy interesante en cuanto a la replicación en virus, y cuestión invariablemente cierta para virus de ADN doble cadena, es que la activación de la replicación del ADN se encuentra vinculada con la regulación de la transcripción genética. Es decir que el ciclo infeccioso normal de un virus de ADN doble cadena está dictado por la función de ciertos genes ‘encendidos’, los cuales dan lugar a la replicación, y que durante la replicación se ‘apagan’, pero se ‘encienden’ otros genes que van a encapsular el material genético sintetizado y así contaminar otras células (Flint *et al.* 2008; Flint *et al.* 2008). Este proceso es llamado en inglés como ‘switch’ genético y se lo encuentra presente en múltiples sistemas biológicos además de virus (Ptashne 2004). Entonces, en virus de ADN doble cadena la replicación representa el momento en el cual cambia el programa de transcripción, de sintetizar proteínas llamadas tempranas, las requeridas como maquinaria de replicación, OBPs virales, entre otras funciones, a sintetizar proteínas tardías, las requeridas para formar los elementos estructurales de la cápside o los elementos proteicos que forman parte del virión.

El reconocimiento de los sitios de ADN específicos para OBPs da identidad al origen de replicación, *ori*, y desencadena la construcción de nuevas progenies virales. Esta función de OBP es cumplida por ejemplo en el virus SV40 por la proteína *Large T antigen* (Kelly 1988), en el herpesvirus asociado a sarcoma de Kaposi por la proteína LANA1 (Komatsu *et al.* 2004), en el virus Epstein-Barr por la proteína EBNA1 (Polvino-Bodnar *et al.* 1992) y en papillomavirus por las proteínas E1 y E2 (Chiang *et al.* 1992). En la mayoría de los casos se descubrió que la unión de las OBPs virales mencionadas a su secuencia específica de ADN está además involucrada en la regulación de la transcripción viral (Cox *et al.* 1988; Polvino-Bodnar y Schaffer 1992; Lim *et al.* 2000; Thierry 2009) y en la segregación cromosómica (Rawlins *et al.* 1985; Komatsu *et al.*

2004; Abbate *et al.* 2006; Feeney *et al.* 2009). En el origen de replicación, la mayoría de los virus tumorales poseen múltiples copias de secuencias de ADN diana para OBPs (Rawlins *et al.* 1985; Titolo *et al.* 2003; Titolo *et al.* 2003; Wong *et al.* 2005; Sanchez *et al.* 2008).

La unión cooperativa (Wyman y Gill 1990) de proteínas reguladoras a dichas secuencias diana múltiples es esencial para una regulación precisa y económica del reclutamiento de la maquinaria de transcripción y/o replicación (Ptashne 2004). Experimentos *in vitro* tales como ensayos de cambio en la movilidad electroforética (EMSA, acrónimo de sus siglas en inglés) muestran que en los casos ensayados de las OBPs virales mencionadas, la unión a múltiples sitios de ADN diana ubicados en orígenes de replicación se desarrolla mediante cooperatividad positiva (Spalholz *et al.* 1988; Tan *et al.* 1994; Summers *et al.* 1996; Weisshart *et al.* 1999; Garber *et al.* 2002; Titolo *et al.* 2003). En el caso de la proteína E2 de HPV, la unión cooperativa a dos sitios en tándem puede ser interpretada como la garantía del *switch* genético vinculado a la replicación, dado que se sabe que la unión a los sitios dispuestos en tándem en forma cooperativa reprime la expresión de las proteínas tempranas E6 y E7, y activa la replicación y la transcripción de proteínas tardías vía la formación de un bucle nucleotídico en el LCR entre los sitios BS1, BS2 y BS4. La termodinámica y cambios estructurales que acompañan la unión cooperativa de proteínas a ADN son claves para el reconocimiento, la torsión y apertura del ADN doble cadena durante su transcripción y replicación, pero no son todavía entendidos con claridad.

Con el objetivo de entender la naturaleza de la interacción entre E2C y ADN, realizamos un análisis detallado del complejo utilizando herramientas bioinformáticas y el contenido de información en las secuencias proteicas y nucleotídicas y usando técnicas biofísicas, proteínas E2C recombinantes y sitios sintéticos de ADN específicos y no-específicos.

Antes de estudiar el rol de la interacción E2C-ADN, nos focalizamos en el proceso de polimerización no-funcional de E2C dado que su desarrollo es reprimido por la presencia de ADN específico. Inducida por temperatura, la polimerización de E2C se desencadena mediante la formación de un núcleo estructurado y finaliza en un paisaje morfológico de oligómeros solubles con propiedades amiloideas.

El estudio computacional de la interacción E2-ADN de todos los tipos de HPV mucosos indica que las proteínas E2 de distintos tipos poseen propiedades de discriminación de ADN

similares. Las diferencias en la interacción E2-ADN se encuentran principalmente en la secuencia de ADN específico, en acuerdo con el análisis de la unión de E2C a diferentes sitios mediante titulaciones isotérmicas de calorimetría. Basado solamente en la secuencia de ADN, profundizamos la descripción de la especificidad de unión de E2C utilizando logos de secuencia (Schneider *et al.* 1990), y revelamos una fuerte selección para un sitio de metilación ubicado dentro de un sitio de unión de E2. Mostramos que la afinidad de E2C a sus sitios diana puede ser predicha en forma adecuada a partir de un alineamiento de sitios naturales, y utilizamos el resultado para recalcular la jerarquía de los sitios para todos los tipos. Fuimos también capaces de identificar características moleculares de la interacción que son significativamente sobre- y subrepresentadas en tipos de alto-, bajo-riesgo y tipos de HPV cutáneos.

Además, presentamos un estudio en solución de la interacción de E2C y ADN de 5 dominios de E2C de los tipos de HPV 11, 16, 31 y 45 en las mismas condiciones fisiológicas. Es por ello que en la presente tesis los resultados expuestos fueron analizados en forma relativa entre las distintas proteínas homólogas y no en forma puntual para cada construcción, como se ha descrito para el dominio hasta el momento. Planeamos medir entonces la estabilidad al equilibrio y cinética de cinco dominios E2C homólogos y por otro lado la capacidad funcional de discriminación de secuencias de ADN y de reconocimiento cooperativo. La desnaturalización al equilibrio y la cinética de desplegado en cloruro de guanidinio mostró que todos los dominios son homodímeros estables que se desnaturalizan vía un mecanismo de dos estados. Medimos la unión a distintos sitios de ADN y confirmamos que el mecanismo de unión para todos los complejos es el mismo en base a una genuina compensación entropico-entálpica.

Nuestros resultados muestran que la inusual estructura de barril-beta de E2C puede acomodarse a numerosas mutaciones manteniendo las propiedades cruciales de conformación y función. Pero la unión cooperativa a sitios adyacentes de ADN mostró que los componentes entálpico-entrópicos de la reacción como la deformación del ADN puede divergir entre distintas homólogas en acuerdo con un fuerte acoplamiento entre la dinámica global de la proteína y el reconocimiento del ADN.

Segunda Sección

Materiales y métodos

- Alors?... dit le directeur.

- Eh bien, voilà!... dit Colin.

- Que savez-vous faire? demanda le directeur.

- J'ai appris des rudiments..., dit Colin.

- Je veux dire, dit le directeur, à quoi passez-vous votre temps?

- Le plus clair de mon temps, dit Colin, je le passe à l'obscurcir.

Boris Vian, L'Écume des jours

1) Reactivos.

- Drogas y equipamiento lógico.

Drogas. Los medios de cultivo utilizados en la expresión de proteínas se prepararon en el laboratorio con extracto de levaduras y tripteína bacteriológica (Laboratorios Britania), con glucosa (Merck S.A.I.C.), y NaCl (Anedra SA). Los tampones químicos utilizados para purificación de proteínas y ensayos biofísicos (espectroscopía de fluorescencia, de absorbancia y dicroísmo circular, ITC, AFM, DLS) fueron preparados en el laboratorio utilizando drogas de alta pureza provenientes de USB Inc., Amresco Inc., Merck S.A.I.C. e ICN Inc., Sigma Aldrich, Fisher Scientific y Promega Inc. Los tampones fueron llevados al pH deseado, titulando con HCl y NaOH (Cicarelli S.A) en un potenciómetro modelo 420A (Orion Instrumentalia). Las drogas fueron pesadas en balanzas digitales, mezcladas y homogeneizadas en agitadores (MicroStirrer, VELP, Scientifica). Todas las diálisis se realizaron en tripas (Spectrum Spectra/Por, Thomas Scientific USA) con Corte de Peso Molecular (MWCO) de 5.5 a 14 kDa y volúmenes de 0.33 a 8 ml/cm.

Equipamiento lógico. Las tablas de datos fueron procesadas en Microsoft Excel y ProFit (v 6.0, QuantumSoft, Suiza). Los ajustes de estabilidad y las estadísticas fueron calculadas en ProFit. Los gráficos de los datos fueron representados en ProFit. Los datos de ITC fueron analizados en SEDPHAT (Houtman *et al.* 2007) y MicroCal Origin. Las secuencias fueron visualizadas y analizadas en programas gratuitos (Tippmann 2004): BioEdit (Bioedit), Clustal X (Larkin *et al.* 2007), Protparam (Gasteiger *et al.* 2003); LogoBar (Perez-Bercoff *et al.* 2006); weblogo (Crooks *et al.* 2004); Two Sample Logo (Vacic *et al.* 2006); DNA Strider (Douglas 1995); Jalview (Clamp *et al.* 2004). Las estructuras cristalográficas y de RMN en PyMOL (Delano) y las mediciones de AFM en WsxM (Horcas *et al.* 2007). La presente tesis fue escrita en el procesador de textos Microsoft Word, fuente Times New Roman, tamaño 12, en una computadora Macintosh iBook PowerPC G4 1.07 GHz, disco rígido de 30 giga octetos (GB) y memoria de acceso aleatorio (RAM) de 1.25 GB bajo el sistema operativo Mac OS X versión 10.4.11, la tesis fue sistemáticamente almacenada en Dropbox.

- Proteínas recombinantes.

Diseño y clonado. Las secuencias de ADN codificando los dominios homólogos de E2C

utilizados en el presente trabajo fueron obtenidas a partir de sus correspondientes genomas. Los genomas de HPV-6, 11, 18 y 45 fueron enviados a nuestro laboratorio generosamente por la Dra Ethel-Michelle de Villiers (HPV6, 11, 18, 16 y 45) y por la Dra Alison McBride (HPV31). Se diseñaron cebadores para cada dominio con el fin de amplificar por PCR la secuencia correspondiente a la región codificante de E2C. El producto de la amplificación fue clonado en el plásmido de expresión ptzU18u. Para ello, los cebadores fueron provistos de secuencias de sitios de corte para las enzimas de restricción NdeI y BamHI, ambos presentes en el vector de interés. Finalmente, todas las construcciones fueron confirmadas por secuenciación nucleotídica y guardadas en la base de cepas del laboratorio a -80 C. El diseño y clonado de la versión monomérica de E2C-16, scE2C se efectuó en trabajos anteriores a esta tesis y se encuentra detallado en (Dellarole 2005; Dellarole *et al.* 2007).

Expresión y purificación. La expresión y purificación de los dominios de E2C y de scE2C se llevó a cabo en tres pasos (Mok *et al.* 1996).

Expresión: los plásmidos ptzU18u clonados con el inserto correspondiente se transformaron en bacterias de expresión *Escherichia coli* BL21(DE3)Plys, siendo la presión de selección los antibióticos cloranfenicol y ampicilina. Se inoculó una colonia de bacterias en 10 ml de medio 2TY con 30 µg/ml de cloranfenicol, 100 µg/ml ampicilina y 1% m/V de glucosa y se incubó 16 horas en agitación a 37°C. Luego, se inoculó el precultivo de 10 ml en 1 l de medio 2TY (16 g peptona, 10 g extracto de levadura y 5 g NaCl) con 30 µg/ml de cloranfenicol y 100 µg/ml ampicilina y se incubó 3-4 horas en agitación a 37°C o hasta obtener una densidad óptica de 0.4-0.6 (DO_{600nm}). Se indujo el cultivo con 0.4 mM de IPTG y se incubó 16 horas en agitación a 37°C. Todos los dominios fueron expresados de esta forma excepto E2C-31 y E2C-6 para los cuales se redujo de 16 a 4 horas el tiempo post inducción con IPTG.

Purificación inicial: se precipitó el cultivo bacteriano por centrifugación a 6000 rpm durante 15 minutos y se resuspendió en 50 ml de tampón químico de lisis (100 mM Tris-HCl pH 6.8, 0.6M NaCl, 1 mM EDTA y 10 mM β-mercaptoetanol). La suspensión se congeló a -20°C en un Falcon de 50 ml contenido en un vaso de precipitado de 500 ml con hielo granizado y etanol. Se descongeló la suspensión y se sometió a sonicación manteniendo la temperatura en 4°C. La sonicación consistió en dos ciclos de tres

sonicados de 10 segundos de duración, potencia 8, con descansos de 30 segundos entre cada sonicado. Luego se separó la fracción soluble de la insoluble mediante 30 minutos de centrifugación a 12000 rpm. Se recuperó la fracción soluble y se precipitó con sulfato de amonio al 80% de saturación (512 g/l) durante 2 horas en agitación a 4°C. Se precipitó la muestra sometida al tratamiento salino (centrifugación a 12000 rpm, 30 minutos) y se resuspendió en una solución amortiguadora (100 mM Tris-HCl pH 6.8, 1 mM EDTA, 0.3 a 0.6 M NaCl y 2 mM 2-mercaptoetanol). Se dializó la resuspensión en tripas de Corte de Peso Molecular de 6-8 kDa (8.0 ml/cm), contra 2 períodos de 1 litro del mismo tampón durante 12 horas a 4°C. Este paso de purificación se realizó de igual manera en todas las proteínas E2C.

Purificación final: la muestra dializada se inyectó a 2 ml/min en una columna de 8 ml de heparina Hyper D (BioSeptra, Villeneuve la Garenne, Francia) equilibrada en el mismo tampón químico a pH 8. En estas condiciones E2C queda retenida en la columna por su afinidad inespecífica a la heparina, característica usual de factores de transcripción. Monitoreando por absorbancia a 280 nm, se eluyeron las proteínas adheridas a la columna aumentando progresivamente la fuerza iónica hasta 2.0 M NaCl en 30 minutos. Se recolectaron fracciones de 5 ml, y aquellas conteniendo E2C con un grado de pureza superior a 85% de acuerdo a gel SDS PAGE 15% de poliacrilamida, fueron juntadas y dializadas en tripas de Corte de Peso Molecular (MWCO) 6-800 kDa (8.0 ml/cm), en solución amortiguadora (10 mM Acetato de Sodio pH 5.6, 0.2 M NaCl, 1mM DTT). El resultado de la diálisis se concentró a 4-6 ml por ultrafiltración en Centriprep-3 (Amicon, Bedford, MA, EEUU). El extracto concentrado se inyectó en una columna de tamiz molecular, Superdex75 (Pharmacia Biotech, Uppsala, Suecia), equilibrada en el mismo tampón químico. Monitoreando por absorbancia a 280 nm, E2C eluye como un único pico (ver Discusión) correspondiente a 18-20 kDa cuyo grado de pureza es superior a 95%, de acuerdo a gel SDS PAGE 15% de poliacrilamida. Finalmente, la muestra se concentró a 4-6 ml por ultrafiltración en Centriprep-3.

El paso por la columna de heparina presentó diferencias entre las E2C. La heparina se utiliza para purificar factores de transcripción dada su alta densidad en cargas negativas y por lo tanto su semejanza con el ADN o ARN. De este modo, la interacción inespecífica de E2C con la heparina es electrostática y entonces sensible a la fuerza iónica gobernada exclusivamente por la

concentración de NaCl. La heparina se equilibró en diferentes tampones químicos variando la concentración de NaCl de acuerdo a la eficiencia de interacción de los dominios E2C: 0.6 M de NaCl para E2C-16 y E2C-18, 0.5 M de NaCl para E2C-11, y E2C-45, 0.4 M de NaCl E2C-6 y 0.3 M de NaCl para E2C-31.

Almacenamiento y control de calidad. Se centrifugaron las proteínas 3 minutos a 12000 rpm y se midió su concentración espectrofotométricamente (absorbancia a 280 nm) utilizando el coeficiente de extinción molar ($\epsilon_{280\text{nm}}$) correspondiente (Pace *et al.* 1995). Como control de calidad de las muestras, se midió directamente el peso molecular de los monómeros con la técnica de ionización suave MALDI-TOF (Microflex Bruker). Además, se corroboró la naturaleza dimerica de cada dominio evaluando el peso molecular de las proteínas en solución. Para ello, se comparó con proteínas estándar la elusión del tamiz molecular y se midió el diámetro hidrodinámico por dispersión de la luz (DLS) (ver Capítulo IV). Finalmente, se analizaron los dominios purificados en cuanto a su estado de plegamiento monitoreando por CD la presencia de estructura secundaria y por fluorescencia de triptófanos la presencia de estructura terciaria en condiciones nativas (N), desnaturalizantes (U) y replegadas (R). scE2C-16 y E2C-16 fueron sometidas a los mismos controles anteriormente (Mok *et al.* 1996; Dellarole *et al.* 2007).

Oligonucleótidos sintéticos.

Procedencia de las secuencias nucleotídicas. Las secuencias de los oligonucleótidos utilizados en la presente tesis consisten por un lado en sitios naturales diana de E2 de HPV tipo 16, los sitios aislados BS1, BS2, BS3 y BS4, en tándem, DBS y la versión mutada de BS3 por inserción de una timidina, BS3* (**Tabla 1**). Por otro lado, se utilizaron los sitios diseñados recto e inespecífico y un sitio diana de E2 papilomavirus bovino tipo 1, BPV-1 (**Tabla 1**). Para mayor claridad, en la **Tabla 1** se muestran únicamente las secuencias en la dirección 5'-3'. Los oligonucleótidos de simple cadena fueron adquiridos comercialmente (Integrated DNA Technologies, Inc Coralville, IA, EEUU) dónde luego de la síntesis fueron purificados por desalado estándar o por HPLC y enviados por correo en estado sólido.

Tabla 1. Secuencias y coeficiente de extinción molar de los oligonucleótidos.

Oligonucleótidos	Secuencias	$\epsilon_{260 \text{ nm}}$ ($\text{mM}^{-1} \cdot \text{cm}^{-1}$)
E2-BS1	AACCGAAACCGGTT	369.3
E2-BS2	AACCGAAATCGGTT	352.4
E2-DBS	AACCGAAATCGGTTGAACCGAAACCGGTT	715.1
E2-DBS-FITC	AACCGAAATCGGTTGAACCGAAACCGGTT	736.1
E2-BS3	AACCCAAAACCGGTT	347.4
E2-BS3*	AACCCAAAATCGGT	352
E2-BS4	AACCGAATTCGGTT	345
Sitio bovino	GACCGACGTCGGTC	332
Sitio recto	AACCGACGTCGGTT	347.2
Sitio inespecífico	ATGGACCTGTCAAG	353.8

Se muestran solamente los nucleótidos involucrados en la unión de E2 y que por lo tanto afectan la constante de disociación, todas las secuencias poseen dos nucleótidos adicionales en sus extremos. En negrita se resaltan los nucleótidos que contactan directamente a E2. Se enseñan todas las secuencias en la dirección directa (5' a 3'). La secuencia de E2-BS3 y de E2-BS3* corresponden a la secuencia reversa-complementaria. El $\epsilon_{260\text{nm}}$ considera las secuencias complementarias. La sigla FITC corresponde a la modificación con isotiocianato de fluoresceína en el extremo 5' mediante un conector de 6 carbonos.

Apareamiento de oligonucleótidos. Los oligonucleótidos de simple cadena se resuspendieron en agua y se cuantificaron espectrofotométricamente en base a su respectivo coeficiente de extinción molar, $\epsilon_{260\text{nm}}$, en una solución de entre 0.2 y 0.5 N de NaOH. Luego se mezclaron las hebras complementarias en cantidades equimolares a una concentración de entre 30 y 200 μM en el tampón químico 10 mM BisTris-HCl pH 7.0, 20 mM NaCl. La mezcla se incubó por 5 min a 95°C y se dejó enfriar progresivamente durante 16 horas hasta alcanzar 25°C. Se confirmó el correcto apareamiento de las hebras complementarias mediante un gel nativo de poliácridamida al 20% (no mostrado). Los oligonucleótidos doble cadena se cuantificaron por absorbancia de acuerdo a su correspondiente $\epsilon_{260\text{nm}}$ (**Tabla 1**) y se almacenaron a -20°C.

2) Métodos.

- Estudios de agregación.

Caracterización amiloidogénica. Muestras de scE2C en una concentración de 10 μM fueron sometidas a 70°C durante 4 minutos en el tampón químico 10 mM Hepes pH 7.4, 50 mM NaCl, 1 mM MgCl_2 , 1 mM DTT. Luego, muestras con y sin tratamiento fueron monitoreadas a

25±0.1°C en los siguientes ensayos:

Dicroísmo Circular. Fueron medidos espectros de CD (Jasco Inc. modelo J-815) en una cubeta de 0.1 cm de paso óptico. Las señales de elipticidad del CD fueron transformadas a elipticidad molar por residuo, calculada según:

$$[\Theta]_{MRW} = \frac{mdeg}{10 \times l \times [\pi] \times (\#ep)} \quad [1]$$

dónde mdeg (miligrados) es la señal observada, l el ancho de la cubeta de medición (en cm), $[\pi]$ la concentración proteica (en molar) y $\#ep$ el número de enlaces peptídicos del dominio analizado.

Tioflavina T. Se equilibraron las muestras 10 minutos a 25°C con 20 μM de 4-(3,6-dimetil-1,3-benzotiasol-3-ium-2-il)-N,N-dimetilanilina clorídrico (Tioflavina T) y se midieron espectros de fluorescencia (Aminco Bowman Series 2 (AB2), Spectronic Co. Rochester, NY, EEUU) excitando a 442 nm y colectando la señal de emisión a 482 nm con un ancho de banda para la excitación y la emisión de 4 nm.

Rojo Congo. Se equilibraron las muestras 10 minutos a 25°C con 5 μM de ácido bencidina diazo-bis-1-naftilamina-4-sulfónico (Rojo Congo) y se midieron espectros de absorbancia (Jasco Inc. modelo V-550).

ANS. Se equilibraron las muestras por 10 minutos a 25°C con 100 μM de 1-anilino-8-naftaleno sulfonato (ANS) y se midieron espectros de fluorescencia excitando las muestras a 370 nm y colectando la emisión a 490 nm con un ancho de banda de 4 nm para la excitación y la emisión.

Al menos tres espectros fueron promediados para todos los casos.

Cinéticas isotérmicas. Las cinéticas fueron registradas en un espectropolarímetro Jasco Inc. modelo J-815 (NikoCota, Japón) siguiendo la señal de scE2C a 225 nm en función del tiempo. La temperatura de la muestra se mantuvo constante mediante un controlador Peltier conectado al soporte de la celda del CD con una fidelidad de ± 0.1°C. En todos los casos, el tampón químico de medición consistió en 10 mM Hepes pH 7.4, 50 mM NaCl, 1 mM MgCl₂, 1 mM DTT. Primero, se llevó a temperatura el tampón químico en la celda del CD durante 5 minutos. Luego, con una jeringa de 250 μl (Hamilton) se inyectó en la cubeta la muestra proteica homogenizando

mediante dos reflujos con la misma jeringa. Con el fin de minimizar variaciones de temperatura al momento de la inyección, el volumen inyectado fue tal que resultara al menos 30 veces menor que el volumen total contenido en la cubeta. El tiempo muerto de medición resultó ser como máximo de 15 segundos.

Cálculo de Parámetros. Las cinéticas isotérmicas fueron analizadas en base a cuatro parámetros: s_i , s_f , t_{50} y VE (**Figura I.7**). Siendo s_i y s_f los valores promedio de la señal de CD a tiempo inicial y final respectivamente, t_{50} el tiempo al 50% de cambio en la señal de CD y VE la velocidad de elongación o mayor pendiente en el cambio de señal de CD. Los parámetros s_i y s_f fueron calculados ajustando el conjunto de datos iniciales o finales a una constante $y = s$. El parámetro t_{50} fue calculado a partir del valor de la señal de CD al 50% del cambio total de señal. El parámetro VE fue calculado ajustando el conjunto de valores de la señal de CD al 33% del cambio total de señal a una recta $y = ax + b$, siendo la pendiente “a”, la VE (**Figura I.7**).

Se registraron cinéticas en las siguientes condiciones:

Presencia de ADN específico. Se incubó 30 minutos scE2C con BS2 en una relación molar 1:1, a una concentración de 10 μM y temperatura de 25°C. Se registró la señal de CD a 225 nm en función del tiempo a una temperatura de 65°C.

Dependencia con la concentración. Se registró la señal espectroscópica de scE2C en un rango de concentraciones de 1 a 25 μM a una temperatura fija de 65°C o 70°C. Las condiciones de 4, 8, 10, 15 y 25 μM fueron medidas en cubetas de cuarzo de 0.1 cm de camino óptico. Las muestras de 1 y 2 μM fueron medidas en cubetas de cuarzo de 0.5 y 1 cm de camino óptico con el fin de optimizar la señal-ruido. Todas las condiciones fueron repetidas hasta tres veces. Luego de explorar diferentes modos de interpretación como de análisis de las curvas, se optó por el modelo y conjunto de parámetros más robusto de Powers y Powers (Powers *et al.* 2006), descrito en el Capítulo I.

Dependencia con la temperatura. Se registró la señal de scE2C a una concentración de 10 μM en nueve condiciones térmicas comprendidas en el rango de 60 a 72.5°C. Todas las condiciones fueron medidas en cubetas de cuarzo de 0.1 cm de paso óptico y fueron repetidas hasta tres veces. Los parámetros t_{50} y VE extraídos de las cinéticas fueron analizados mediante la ecuación de Arrhenius, llamada así por el descubridor del

conocido efecto invernadero, Svante Arrhenius:

$$E_a = -RT \ln\left(\frac{k}{A}\right), \quad [2]$$

siendo $k(t)$ la constante cinética, A el factor pre-exponencial, E_a la energía de activación, R la constante universal de los gases y T la temperatura absoluta (K°). Para ser usada como modelo de regresión lineal en gráficos de pseudo Arrhenius, entre las variables medidas y T^{-1} , esta ecuación puede ser reescrita como,

$$\ln(VE) = \ln(A) - \frac{E_a}{RT}, \text{ o bien } \ln(t_{50}) = \ln(A) - \frac{E_a}{RT} \quad [3]$$

En todos los casos, la señal espectroscópica fue registrada hasta al menos tres veces el tiempo t_{50} y normalizada calculando el correspondiente MRW, ecuación [1].

- *Análisis Morfológico.*

Dispersión dinámica de la luz. El diámetro hidrodinámico de scE2C fue medido en un equipo de dispersión dinámica de la luz (DLS), Malvern Instruments. Por un lado, se realizaron mediciones cada $3^\circ C$ en el intervalo de $25^\circ C$ a $85^\circ C$ y calentando a una velocidad de $3^\circ C/\text{minuto}$. El calentamiento se realizó mediante el Peltier acoplado al soporte provisto por el DLS. Por otro lado, se registraron por lo menos diez mediciones a $25^\circ C$ de muestras de scE2C sin tratar o calentadas a $85^\circ C$. Las mediciones fueron promediadas mediante el operador lógico provisto por el equipo, *Dispersion Technology Software* (DTS) versión 4.2 Malvern Instruments y cuantificadas considerando una distribución normal de los datos y ajustando a la función gaussiana en dónde cada pico se describe por,

$$A \exp\left(\frac{-(x - x_0)^2}{2\sigma^2}\right) \quad [4]$$

dónde A es la amplitud, x_0 la media aritmética y σ el desvío estándar. Se tomó 2σ como el error de medición, intervalo de confianza 95%. Los experimentos se llevaron a cabo a una concentración de scE2C de $16.5 \mu M$ en el tampón químico 10 mM Hepes pH 7.4, 50 mM NaCl, 1 mM $MgCl_2$, 1 mM DTT.

Microscopio de fuerza atómica. Las muestras de scE2C fueron tratadas a $70^\circ C$ durante 4 minutos en el tampón químico 10 mM Hepes pH 7.4, 50 mM NaCl, 1 mM $MgCl_2$

previamente purificado con filtros de 0.22 μm . La solución amortiguadora con o sin scE2C y sometida o no al tratamiento térmico fue depositada y fijada en discos de mica. Muestras fijadas de 2, 10 y 20 ng de scE2C fueron medidas en el servicio de microscopia de fuerza atómica a cargo de Silvio Ludueña y coordinado por la Dra Lía I. Pietrasanta del Centro de Microscopías Avanzadas (CMA) ubicado en el Pabellón 1 de la Ciudad Universitaria, FCEN, UBA. Las imágenes fueron corregidas e interpretadas en el laboratorio mediante el operador lógico gratuito WsxM v. 3.0 (Horcas *et al.* 2007). Las morfologías identificadas se analizaron en base a tres parámetros: el ancho, correspondiente al 50% del máximo de intensidad (ancho total máximo medio o FWHM), la altura, correspondiente al máximo valor en el eje z, y en el caso de fibras tipo gusano se midió además la longitud de contorno, correspondiente al largo de los mismos. Los parámetros de las morfologías tipo bola fueron analizadas estadísticamente mediante el operador lógico proFit, considerando una distribución normal de los datos y ajustando a la función gaussiana [4].

Microscopía electrónica: tinción negativa y criomicroscopía. Las muestras de scE2C fueron tratadas a 70°C durante 4 minutos en el tampón químico 10 mM Hepes pH 7.4, 50 mM NaCl, previamente purificado con filtros de 0.22 μm y enviadas al laboratorio de Estructura Molecular y Función, Departamento de Ciencias Biológicas, Colegio Imperial de Londres, Inglaterra, co-dirigido por el Profesor Paul Freemont y el Profesor Xiaodong Zhang. La tinción negativa se llevó a cabo incubando las muestras 1 minuto para su absorción en una grilla Formvar (formal polivinilo) recubierta de carbon y teñido en acetato de uranilo 2% pH 4.5, a una concentración de agregados solubles de scE2C de 0.05 mg/ml y la criomicroscopía se llevó a una concentración de agregados solubles de scE2C de 0.5 mg/ml (JEOL 2200FS electron microscope). El procesamiento de imágenes, selección de patrones y reconstrucción tridimensional a partir de 16000 partículas digitalizadas (Nikon Coolscan 9000 ED), fue llevado a cabo en el laboratorio de Estructura Molecular y Función por la amable colaboración con la Dra Cecilia Bebeacua mediante el programa analítico IMAGIC (van Heel *et al.* 1996) y el análisis estadístico de variables múltiples (van Heel 1984).

- **Bioinformática.**

Base de datos de los sitios de unión de E2. Todos los genomas del *genus* alfa fueron obtenidos del Comite Internacional Taxonómico de Base de datos Viral (ICTV, acrónimo de sus siglas en inglés) (Büchen-Osmond 2003), siendo la identificación taxonómica (ID): 151340, detallado en la tabla suplementaria de (Sanchez *et al.* 2008). Los genomas presentes con más de una entrada fueron examinados con el fin de verificar si son o no redundantes mediante la aplicación disponible en internet *BLAST 2 sequences* (Tatusova *et al.* 1999). En la mayoría de los casos, el punto de partida de la secuencia genómica cae dentro de la región reguladora larga (LCR, acrónimo de sus siglas en inglés), partiendo el LCR en dos porciones, una al comienzo y otra al final de la secuencia. Se alinearon todos los genomas utilizando ClustalW (Chenna *et al.* 2003) y se extrajeron las dos secuencias parciales del LCR tomando el comienzo del gen E6 y el final del gen L1 como referencia. Se unieron las dos porciones para obtener el LCR completo, el cual fue despojado de inserciones y vuelto a alinear. Finalmente, los sitios de unión de E2 fueron extraídos y alineados manualmente. Para la manipulación de las secuencias se utilizaron los editores de alineamiento BioEdit v7.0.8 (Tom Hall, Ibis Biosciences) y Jalview v2.2.1 (Clamp *et al.* 2004).

Logos de secuencia. Los logos de secuencia fueron generados con WebLogo (Schneider y Stephens 1990; Crooks *et al.* 2004) y con las secuencias de ADN o proteínas alineadas. La altura de las letras apiladas en una posición i fue calculada según:

$$R_{\text{secuencia}}(i) = \log_2(s) + \sum f(b,i) \log_2(f(b,i)) - \frac{s-1}{2 \cdot \ln(2) \cdot n} \quad [5]$$

dónde s es el número de símbolos (4 para el ADN y 20 para proteínas) y $f(b,i)$ son las fracciones de cada base o amino ácido en la posición i . El tercer término es una corrección para muestras pequeñas, dónde n es el número de secuencias en el alineamiento. El valor máximo de $R_{\text{secuencia}}$ es 2 para ADN y 4,32 para proteínas, siendo el mínimo cero para ambos casos. La altura de cada letra dentro de cada apilamiento es proporcional a su abundancia:

$$\text{Altura}(b,i) = f(b,i) \cdot R_{\text{secuencia}}(i) \quad [6]$$

El logo de dos muestras, *Two Sample Logo* (Vacic *et al.* 2006), fue utilizado para comparar el logo medido *in vivo* contra el logo medido *in vitro*. La aplicación toma como entrada un alineamiento de referencia y un alineamiento de contraste e identifica las posiciones que se encuentran enriquecidas o empobrecidas en una determinada base. El alineamiento de sitios *in vitro* fue generado utilizando la frecuencia de bases reportadas en (Newhouse y Silverstein

2001), y fue utilizado como alineamiento de contraste. Se utilizó el alineamiento de sitios *in vivo* como muestra. Se utilizó el test binomial y un valor p de 0.05 de corte para identificar las diferencias entre los alineamientos.

Predicción computacional para las energías libres de unión. Se utilizó la teoría de Berg y von Hippel (Berg *et al.* 1987; Berg *et al.* 1988; Berg *et al.* 1988) y el alineamiento de los sitios naturales de E2 para predecir la energía de E2 a sus sitios de unión. Todos los cálculos fueron realizados mediante archivos de órdenes (*scripts*) en el lenguaje de programación Perl, hechos para la ocasión y con ProFit. Se utilizaron las secuencias directas de E2-BS1 y E2-BS2 y la secuencia reversa complementaria de E2-BS3 y E2-BS4 de forma tal de alinear todos los sitios en la misma orientación (ver resultados). La teoría asume que la única presión de selección en los sitios de unión naturales presentan una energía de unión por debajo de una ventana dictada por la cantidad de proteína libre en la célula y los niveles de unión requeridos. Las posiciones 4 y 5 para la interacción E2-BS fueron excluidas del cálculo de energía libre de unión dado la presión de selección por un sitio de metilación en esas bases. Realizar el cálculo con las frecuencias de la selección *in vitro* (Newhouse y Silverstein 2001) no cambia los resultados significativamente (datos no mostrados), confirmando que la selección para la unión es la mayor presión de selección en los sitios naturales de unión de E2. Una segunda suposición de la teoría es que los pares de base evolucionan en forma independiente. Se utilizó la aplicación Enologos (Workman *et al.* 2005) para verificar que las correlaciones entre las frecuencias de los pares de base en los sitios de unión naturales de E2 fueran muy débiles o inexistentes (datos no mostrados).

El ruido estadístico esperado en la correlación experimental entre la energía libre de unión calculada es de alrededor de 1 kcal/mol (Berg y von Hippel 1987; Berg y von Hippel 1988; Berg y von Hippel 1988). Entonces, se eligieron conjuntos de datos que abarquen por lo menos 2 kcal/mol de forma tal de poder observar una correlación. Cada conjunto de datos experimentales fue medido bajo diferentes condiciones de tampón químico (Alexander y Phelps 1996; Hines *et al.* 1998; Kim *et al.* 2000; Newhouse y Silverstein 2001; Zhang *et al.* 2004). Dado que los cambios en las condiciones de medición influyen en la capacidad de discriminación de secuencia de la proteína E2 (Blakaj *et al.* 2006), se realizó una correlación por separado para cada set de datos. La pseudo-temperatura T^* se relaciona con la fuerza de la presión de selección *in vivo* para la afinidad. En nuestros cálculos, fue arbitrariamente fijada a 298 K. Por lo tanto, las

pendientes en la **Figura II.4** son equivalentes al cociente de la temperatura experimental y la pseudo-temperatura. Estas pendientes varían más de dos órdenes, probablemente de acuerdo a variaciones en la selectividad de secuencia del dominio de E2C debido a las condiciones de tampón químico (Blakaj *et al.* 2006).

También se aplicó la teoría de Berg-von Hippel para buscar sitios de unión no reportados previamente en el LCR de los alfapapilomavirus. Se calculó la energía de unión para todos los sitios posibles de 14 bases usando la frecuencia de bases del alineamiento de los sitios naturales de unión de E2. En algunos casos, la base presente en un sitio de unión putativo no se encontraba en su posición correspondiente en el alineamiento de sitios naturales. Sin embargo, la contribución de dicha base a la energía de unión no puede ser calculada en forma sencilla, se puede asumirla como altamente perjudicial para la unión. Es posible entonces postular que los sitios con bases no presentes en el alineamiento natural no eran sitios putativos de unión de E2. Se consideró una secuencia como sitio de unión putativo de E2 si su unión predicha era menor que 4kcal/mol. Esta consideración incluyó al 95 % de los sitios de unión de E2 conocidos.

Agrupación de las energías de unión predichas. Se utilizó el algoritmo de media-k (MacQueen 1967) para agrupar los tipos alfapapilomavirus de acuerdo a las energías de unión predichas. Los tipos virales fueron definidos como puntos i en un espacio de cuatro dimensiones utilizando las energías de unión predichas. El algoritmo utiliza como entrada el número de grupos, j . Primero, define un centro en el espacio de cuatro dimensiones para cada categoría, centroide. Luego, se definen los grupos asociando cada punto i al centroide más cercano. Luego, nuevos centroides j son calculados como centros de masas del grupo. La asociación de cada punto y el cálculo de nuevos centroides es repetida hasta que los centroides no se mueven más. Este algoritmo minimiza la suma del cuadrado de las distancias J entre los datos i y el centroide j , c_j :

$$J = \sum_i \sum_j \|x_i - c_j\|^2 \quad [7]$$

Este procedimiento converge en un mínimo que depende significativamente de los centros de grupo elegidos al azar. De forma de obtener un mínimo global, se corrió el procedimiento 1000 veces hasta que convergiera y se guardó la solución con el menor valor J . El algoritmo fue corrido utilizando el programa R (*The R project for statistical computing*, <http://www.r-project.org>).

Estudios de asociación. Se analizó la asociación entre los genotipos de los tipos virales y los

fenotipos, mediante un ensayo de distribución hipergeométrica (Rivals *et al.* 2007). Tenemos una población total de 72 tipos virales, de los cuáles un número $x \leq 72$ goza una determinada propiedad epidemiológica, y $z \leq (x,y)$ de ellos presenta una propiedad epidemiológica de interés. Se usó la función hipergeométrica para calcular la probabilidad de tener z aciertos sin reemplazo y objetos de una población total de 72, siendo x la población de aciertos. En el caso de asociación positiva el valor-p es definido como la suma de probabilidades de tener z o menos aciertos. Una asociación positiva o negativa es reportada si el valor-p es menor que el elegido como corte, 0.05. Todos los cálculos usando la función hipergeométrica fueron implementados en MS Excel.

- Calorimetría y espectroscopia para complejos binarios.

Calorimetría de complejos binarios. Los experimentos de calorimetría fueron realizados en un calorímetro VP-ITC, Microcal Inc. Las muestras de E2C a una concentración de entre 8 y 11 μM fueron tratadas con 10 mM DTT durante una hora a 25°C. Luego, fueron dializadas intensivamente en tripas de MCWO de 6-8 kDa en el tampón químico correspondiente. Las muestras de oligonucleótidos a una concentración de entre 80 y 110 μM fueron dializadas intensivamente en tripas de MCWO de 3.5 kDa junto a las muestras proteicas. El resultado de la diálisis fue centrifugado a 12000 rpm durante 5 minutos, sometido a vacío durante 5 minutos, ThermoVac Inc, y cuantificado. Las celdas del calorímetro fueron cargadas con 1.8 ml de muestra y el inyector automático con 300 μl de ligando. Las inyecciones de oligonucleótidos a la muestra proteica se realizaron en forma automática titulando de entre 6 y 10 μl , salvo la primer inyección de 2 μl , y manteniendo la rotación de la jeringa en 290 rpm. Las titulaciones se realizaron más allá de niveles de saturación con el objetivo de establecer el calor de dilución del ligando. El operador lógico Microcal Origin versión 5.0 provisto por el VP-ITC fue utilizado para integrar los calores medidos y restar el calor de dilución. Los ajustes globales de los datos calorimétricos fueron realizados en el operador lógico gratuito SEDPHAT (Houtman *et al.* 2007). La ‘sigmoidicidad’ de las titulaciones fue evaluada en base al ‘valor C’ (Wiseman *et al.* 1989):

$$C = nK[\pi] \quad [8]$$

siendo n el número de sitios por ligando, K la constante de afinidad y $[\pi]$ la concentración de la muestra. Valores de c entre 1 y 1000 validan la constante de afinidad medida (Wiseman *et al.* 1989).

Dependencia del pH. El tampón químico de tres componentes (Ellis *et al.* 1982), 100 mM Tris, 50 mM MES, 50 mM acetato de sodio, 200 mM NaCl y 0.2 mM DTT, fue preparado disolviendo los pesos apropiados de cada componente en agua. El DTT fue agregado desde una solución madre 1 M. La solución resultante presentó un pH de 8.3 – 8.5, el cual fue llevado al pH deseado titulando con HCl o NaOH. El pH fue medido en un potenciómetro Orion Instrumentalia, modelo 420A. Se utilizó el operador lógico VP-ITC provisto por Microcal Origin para el ajuste de los datos de ITC.

Se ajustaron los datos de las afinidades a un modelo de dos residuos titulantes (Baker *et al.* 1996) en función del pH:

$$\Delta G_{unión(pH)} = \Delta G_{unión}(pH14) + 2RT \ln \left(\frac{1 + 10^{pK_{libre}^A - pH}}{1 + 10^{pK_{complejo}^A - pH}} \right) + 2RT \ln \left(\frac{1 + 10^{pK_{libre}^B - pH}}{1 + 10^{pK_{complejo}^B - pH}} \right) \quad [9]$$

donde A y B son los dos grupos titulantes y el factor 2 proviene de la naturaleza homodimérica de E2C.

Dependencia de la fuerza iónica. La relación entre la constante de afinidad K y la fuerza iónica, FI , ha sido interpretada de acuerdo a las estequiometrías relativas de cationes, aniones y moléculas de agua involucradas en la formación de complejos proteína-ADN (Record *et al.* 1978; Record *et al.* 1991; Ha *et al.* 1992; O'Brien *et al.* 1998), de acuerdo a la siguiente ecuación simplificada,

$$\log K_D = \log K_{ref} - SK_{obs} \cdot \log[FI] + 0.016 \cdot B[FI] \quad [10]$$

donde SK_{obs} corresponde al número neto de iones desplazados (valores positivos) o reclutados (valores negativos), B al número total de moléculas de agua desplazadas durante la formación del complejo, y K_{ref} es la constante de unión hipotética a 1 M de FI . A bajas FI (< 0.5 M) la actividad del agua se la considera nula, (Record *et al.* 1991), y

$$\log K_D = \log K_{ref} - SK_{obs} \cdot \log[FI] \quad [11]$$

Finalmente, de acuerdo a la naturaleza poli-electrolítica del ADN se considera que SK_{obs} posee una mayor contribución catiónica (Record *et al.* 1991).

Espectroscopia de complejos proteína-ADN. Los espectros de dicroísmo circular y absorbancia fueron registrados en el equipo Jasco (Japón) J815 en una celda de cuarzo de 0.1 cm de paso óptico. Todos los complejos fueron analizados a una concentración de 15 μ M. Todos los complejos fueron incubados por 30 minutos a 25°C y centrifugados 15 minutos a 13000 rpm

antes de medirlos. Diez acumulaciones fueron registradas y promediadas para cada medición. El espectro del tampón químico fue registrado y sustraído en los espectros medidos con muestras. El análisis de los datos fue realizado en el operador lógico ProFit.

- Estudios de Estabilidad.

Desnaturalización al equilibrio. Las desnaturalizaciones químicas al equilibrio de las cinco proteínas homólogas E2C fueron inducidas por cloruro de guanidinio en 50 mM fosfato de sodio pH 7; 1 mM de DTT, variando de 0 a 6 M la concentración de Gdm.Cl, a $25 \pm 0.1^\circ\text{C}$ y el estado de plegamiento al equilibrio fue monitoreado por fluorescencia y dicroísmo circular como se detalla a continuación:

Fluorescencia. Las muestras proteicas fueron excitadas a 280 nm (4 nm de paso de óptico) y el espectro de emisión fue registrado de 295 a 450 nm (8 nm de paso de óptico) a una velocidad de barrido de 5 nm por segundo, acumulando al menos 2 mediciones por muestra. La concentración de E2C final fue de 10 μM y el volumen final de medición de 300 μl , utilizando una cubeta de cuarzo de 0.4 cm de paso óptico. El centro de masa del espectro de emisión (longitud de onda de emisión promedio) fue calculado de acuerdo a:

$$\frac{\sum \lambda I(\lambda)}{\sum I(\lambda)} \quad [12]$$

donde λ es la longitud de onda de emisión e $I(\lambda)$ es la intensidad de fluorescencia a una longitud de onda λ ; y graficado en función de la concentración de Gdm.Cl.

Dicroísmo circular. Las muestras proteicas fueron analizadas midiendo el espectro de CD registrado de 215 a 260 nm (2 nm de paso de óptico) a una velocidad de barrido de 100 nm por minuto, acumulando 5 mediciones por muestra. La concentración de E2C final fue de 10 μM y el volumen final de medición de 300 μl , utilizando una cubeta de cuarzo de 0.1 cm de paso óptico. La señal de CD fue normalizada a MRW, ecuación [1], y luego la señal de CD a 225 nm de longitud de onda fue extraída de los datos promediando los datos de 225 ± 2 nm y graficada en función de la concentración de Gdm.Cl.

Las muestras proteicas fueron incubadas 16 horas en presencia de concentraciones crecientes de Gdm.Cl, monitoreadas por fluorescencia y dicroísmo circular, incubadas 29 horas más y nuevamente monitoreadas por fluorescencia y dicroísmo circular. En el modelo de

disociación de dos estados para dímeros, $N_2 \leftrightarrow 2D$:

$$K_D = e^{\left(\frac{-\Delta G_{eq}}{RT}\right)} = \frac{[D]^2}{[N_2]} \quad [13]$$

donde K_D es la constante de desnaturalización, ΔG_{eq} la energía libre de desnaturalización al equilibrio, N_2 es el estado nativo del dímero y D estado desnaturalizado monomérico. La fracción del estado desnaturalizado se calculó de la siguiente forma:

$$F_D = \frac{(I_N - I_{obs})}{(I_N - I_U)} = \frac{[D]}{(2[N_2] + [D])} \quad [14]$$

donde I_{obs} es la intensidad observada e I_N y I_U son las intensidades de los estados plegados y desplegados descritas por un polinomio de grado 1. El modelo de desnaturalización de dos estados, en el cual el evento de desnaturalización y de disociación se encuentran acoplados (Mok *et al.* 1996), fue ajustado a los datos obtenidos tanto a las 16 horas como a las 45 horas de incubación mediante el operador lógico ProFit. En ambos tiempos de medición para las cinco proteínas E2C ensayadas, el ajuste del modelo a los datos de CD y fluorescencia produjo los mismos parámetros garantizando el estado en equilibrio de las muestras. En todos los casos se asumió que la energía libre de desnaturalización al equilibrio ΔG_{eq} varía linealmente con la concentración del agente desnaturalizante, cuya constante de proporcionalidad es m (Tanford 1968; Greene *et al.* 1974; Pace 1986; Santoro *et al.* 1988; Mok *et al.* 1996). La concentración de desnaturalizante en la cual la mitad de moléculas proteicas se encuentra desnaturalizada $[D]_{50\%}$, fue calculada como se describió anteriormente (Santoro y Bolen 1988; Mok *et al.* 1996).

Finalmente, con el objetivo de garantizar la reversibilidad de la reacción de desnaturalización se realizaron ensayos de replegamiento por dilución de muestras desnaturalizadas. En todos los casos la reacción de desnaturalización química resultó ser reversible en base a la señal de CD y fluorescencia indicando presencia de estructura secundaria y terciaria en las muestras replegadas.

Cinéticas de desnaturalización. Las desnaturalizaciones químicas de las cinco proteínas homólogas E2C fueron inducidas por cloruro de guanidinio en 50 mM fosfato de sodio pH 7; 1 mM de DTT, variando de 3 a 8 M la concentración de Gdm.Cl, a $25 \pm 0.1^\circ\text{C}$ y las cinéticas de la reacción fueron registradas por fluorescencia y dicroísmo circular como se detalla a continuación:

Fluorescencia. Se registró la fluorescencia emitida a 339 nm (5 nm de paso de óptico) de muestras excitadas a 280 nm (5 nm de paso óptico). La concentración de E2C final fue de 0.5 μM y el volumen final de medición de 700 μl , se utilizó una cubeta de cuarzo de 1 cm de paso óptico.

Dicroísmo circular. Se registró la señal de CD a 225 nm (10 nm de paso óptico). La concentración de E2C final fue de 1 μM y el volumen final de medición de 500 μl , utilizando una cubeta de cuarzo de 0.2 cm de paso óptico.

El tampón químico se dejó equilibrar en el equipo de medición entre 130 y 500 segundos en reposo, asegurando la estabilización de la temperatura y entonces el equilibrio en la señal de medición. El agregado de la proteína, en volúmenes de 14 μl en los ensayos de fluorescencia y 10 μl para CD, al tampón químico equilibrado se realizó sin sacar la cubeta del equipo de medición e inyectando la muestra con una jeringa Hamilton de 500 μl , minimizando así variaciones de temperatura. El tiempo muerto de la mezcla en todos los casos fue menor o igual a 10 segundos. De acuerdo con la velocidad de las cinéticas se registraron de uno a diez datos cada cinco segundos. El tiempo total de medición se extendió en todos los casos por lo menos tres veces el tiempo a la mitad de caída de señal. Todas las cinéticas fueron ajustadas a una exponencial simple salvo para las cinéticas medidas con E2C-45 que fueron ajustadas a una exponencial doble. Se corroboró experimentalmente que tanto las mediciones monitoreadas por distintas señales como las mediciones realizadas a distinta concentración generaran los mismos resultados, como era de esperar para cinéticas de desnaturalización química y para un proceso cooperativo entre estructura secundaria y terciaria. Finalmente, los datos de las cinéticas de desnaturalización fueron ajustados a un modelo de dos estados simple, $\text{N}_2 \leftrightarrow 2\text{D}$ (Prat Gay *et al.* 2005). En todos los casos se asumió que las energías de activación varían linealmente con la concentración del agente químico desnaturalizante, con una constante de proporcionalidad, m (Prat Gay *et al.* 2005).

- Calorimetría y espectroscopia para complejos ternarios.

Calorimetría de complejos ternarios. Los experimentos de calorimetría de complejos ternarios fueron realizados como se describió para los experimentos de complejos binarios. En titulaciones directas del complejo ternario, el sitio de ADN DBS (**Tabla 1**) a una concentración de 32.9 μM a 40.5 μM fue titulado en la celda del VP-ITC conteniendo la proteína E2C-11, E2C-16 o E2C-18

en una concentración de 4.5 μM a 4.9 μM . En titulaciones reversas del complejo ternario, E2C en un rango de concentraciones de 41.2 μM a 48.1 μM fue titulado en la celda del VP-ITC conteniendo el oligonucleótido DBS en concentraciones de 3.3 μM a 4.4 μM . Los ajustes globales a un modelo de unión de dos sitios de los datos obtenidos de titulaciones directas y reversas para cada complejo ternario fueron llevadas a cabo mediante el operador lógico gratuito SEDPHAT (Houtman *et al.* 2007).

Estequiometrías por fluorescencia, anisotropía y ensayo de retardo en la movilidad electroforética. Los ensayos de unión a ADN medidos por fluorescencia se realizaron en un espectrofluorímetro Aminco Bowman Series 2. Cuando el fluoróforo fue FITC, se excitó a 490 nm y se colectó la emisión a 520 nm, con un ancho de banda de 4nm para la excitación y 8. Los datos de intensidad de fluorescencia se corrigieron linealmente según la dilución. La intensidad de fluorescencia y anisotropía se realizaron en una situación estequiométrica, es decir, en dichas condiciones experimentales, la concentración fija de proteína o ADN se encuentran más de un orden por encima de la constante de disociación K_D de los sitios individuales medidos mediante la respectiva técnica (no mostrado). Las medidas de anisotropía de fluorescencia se realizaron utilizando polarizadores lineales y colectando las intensidades verticales y horizontales a las mismas longitudes de onda antes mencionadas. Se promediaron al menos 3 mediciones a 0° y 90° y se calculó la anisotropía (r) según (Lakowicz 1999):

$$r = \frac{(I_{vv} - G \times I_{vh})}{(I_{vv} + 2G \times I_{vh})} \quad [15]$$

donde I es la intensidad de fluorescencia, arriba de excitación y abajo de emisión, los sufijos v y h se refieren a polarización lineal vertical y horizontal respectivamente y el factor de corrección G esta definido como $G = \frac{I_{hv}}{I_{hh}}$. Dado el pequeño cambio de rendimiento cuántico del motivo FITC asociado a la unión de ADN (<15%), no fue necesario realizar correcciones de intensidad sobre los valores anisotrópicos (Lakowicz 1999). Las condiciones de medición del ensayo fueron fijadas a 25 mM BisTris-HCl pH 7.0, 200 mM NaCl y 1 mM DTT y se mantuvo constante la temperatura a 25 ± 0.1 $^\circ\text{C}$ utilizando un baño térmico acoplado al fluorímetro. El ensayo de retardo de la movilidad electroforética (EMSA) o “gel shift” se llevó a cabo en condiciones de baja fuerza iónica (Ausbel *et al.* 1997). Las mezclas de reacción fueron incubadas en las condiciones estándar de unión, por 30 min a 25°C en un volumen final de 15 μl . Las mezclas se

sembraron en un gel de poliacrilamida al 6% conteniendo 2,5% de glicerol, 25 mM Tris-HCl pH 8.3, 190 mM Glicina, 1 mM DTT. El gel se corrió entre 5-7mA durante 2 horas en cámara fría. Las bandas de oligonucleótido fluoresceinado fueron escaneadas en un STORM 840 Imaging (Molecular Dynamics) y cuantificadas en el programa Image Quant v1.1.

Cinéticas. Las cinéticas de los cambios estructurales fueron registradas en un espectrofotómetro SX18-MV Applied Photophysics. La absorbancia fue medida a 270 nm (4 nm ancho de banda). Un volumen de 2 μ M de oligonucleótido DBS fue mezclado con un volumen de proteína a una relación proteína:ADN de 2:1, 4:1 y 8:1.

Análisis de poblaciones. Las poblaciones fraccionales de DBS libre, DBS unido a un dominio E2C y del complejo ternario fueron calculadas en base a la concentración de E2C usando los valores medidos para K_{BS1} , K_{BS2} y K_{coop} de la siguiente manera:

Fracción de DBS libre:

$$\frac{1}{1 + K_{BS1}[E2C] + K_{BS2}[E2C] + K_{BS1}K_{BS2}K_{coop}[E2C]^2} \quad [16]$$

Fracción de un sitio unido:

$$\frac{K_{BS1}[E2C] + K_{BS2}[E2C]}{1 + K_{BS1}[E2C] + K_{BS2}[E2C] + K_{BS1}K_{BS2}K_{coop}[E2C]^2} \quad [17]$$

Fracción de dos sitios unidos:

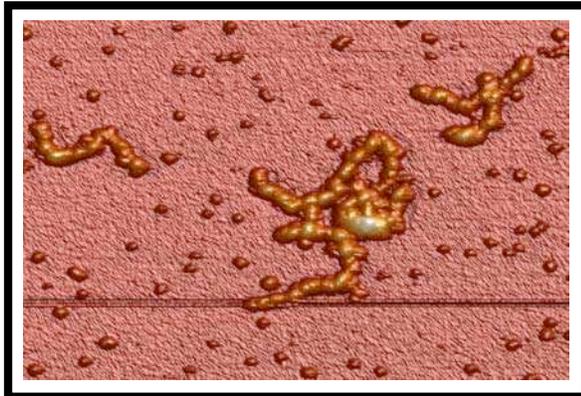
$$\frac{K_{BS1}K_{BS2}K_{coop}[E2C]^2}{1 + K_{BS1}[E2C] + K_{BS2}[E2C] + K_{BS1}K_{BS2}K_{coop}[E2C]^2} \quad [18]$$

Tercera Sección

Resultados

Capítulo I

“Mecanismo de formación de protofibras amiloides en E2C”



*Only rarely is there an opportunity to participate in a molecular happening,
you are going to have that opportunity.*

Paul Berg

Estados de agregación del dominio de unión a ADN E2C.

La propensión de E2C de formar agregados desordenados, es decir precipitados visibles ($> 1 \mu\text{m}$), no-estructurados e irreversibles, es favorecida por un lado por condiciones oxidantes, como las altas temperaturas y pH alcalino (Ferreiro 2003; Nadra 2005). Por otro lado, por interactuar con otras moléculas en condiciones de baja fuerza iónica ($< 0.2 \text{ M}$) como es el caso de hetero-oligomerizaciones entre E2C y ADN específico (Ferreiro *et al.* 2000), y E2C y la proteína E7 (Smal *et al.* 2009). Para E2C, los agregados amorfos son esperables en dichas condiciones ya que un medio oxidante fomenta la unión covalente entre cisteínas expuestas al solvente y la baja fuerza iónica induce interacciones inespecíficas, favoreciendo que las moléculas se acerquen entre sí y agreguen. Sin embargo, cabe destacar que en todos los eventos de agregación desordenada se advirtieron signos de intermediarios solubles. En el caso de co-agregados E2C-ADN, se observaron intermediarios solubles correspondientes a complejos de estequiometría de 2:1 (Ferreiro *et al.* 2000), siendo el segundo evento de unión inespecífico (**ver Figura III.1**). Además, en el caso de co-agregados E2C-E7, se reportó que previo al agregado se pueblan especies solubles presentando características estructurales de lámina beta según lo reportado por el análisis de dicroísmo circular (CD, acrónimo de sus siglas en inglés), de los mismos (Smal *et al.* 2009).

El estudio *in vitro* de intermediarios estructurados suele ser enmascarado por la aparición de agregados desordenados. La formación de agregados estructurados requiere que una parte significativa de la población proteica se encuentre en estado parcialmente desplegado (Kelly 1998; Chiti *et al.* 1999; Ramirez-Alvarado *et al.* 2000), situación también propensa a la formación de agregados desordenados (Dobson *et al.* 1999). La inducción de estados parcialmente desplegados desde estados plegados en proteínas globulares se lleva a cabo empleando condiciones de entorno extremas: pH bajo, temperatura y presión altas o presencia de solventes orgánicos (Marcon *et al.* 2006). Un solvente orgánico muy utilizado es el trifluoroetanol (TFE), el cual, a bajas concentraciones, aumenta la solubilidad de grupos hidrofóbicos (Buck 1998) modulando lenta y progresivamente la estructura de proteínas globulares. Recientemente, se estudió para la proteína E2C el efecto del TFE (Wetzler *et al.* 2007). Dicha perturbación suave en el entorno químico de E2C promueve la formación de agregados estructurados de tipo amiloide (Wetzler *et al.* 2007). Este mismo comportamiento

frente al TFE, se observó tanto en proteínas monoméricas de topología tipo barril *beta* (Srisailam *et al.* 2003; Pallares *et al.* 2004), como en proteínas de diferente topología (MacPhee *et al.* 2000). La caracterización morfológica de los agregados de E2C fue realizada por microscopía electrónica de tinción negativa y presentó tres tipos de especies morfológicamente diferentes: anillos de entre 5 y 10 nm de diámetro, fibras flexibles (gusanos) de 7-8 nm de ancho y 30 nm de largo y fibras rígidas de hasta 300 nm de largo (Wetzler *et al.* 2007). Finalmente se observó que E2C unido a ADN blanco inhibe la formación de estas estructuras amiloideas (Wetzler *et al.* 2007). A nivel experimental, la utilización de TFE como agente promotor de oligomerización presenta ciertas restricciones, como ser la posibilidad de variar la temperatura o la caracterización morfológica por métodos espectroscópicos y por AFM. A la hora de caracterizar los mecanismos moleculares en la formación de agregados amiloideos, resulta clave desarrollar herramientas que permitan explorar diferentes condiciones y sean relevantes en entornos compatibles con el medio celular (Marcon *et al.* 2006).

Como objeto de mi tesina de licenciatura, construimos mediante ingeniería genética una versión monomerizada o ‘simple cadena’ de E2C, scE2C (Dellarole 2005; Dellarole *et al.* 2007). El diseño racional del conector garantizó que las características básicas de topología, estructura y función de E2C se mantuvieran intactas. En otras palabras, la naturaleza aminoacídica del conector, compuesto por residuos hidrofílicos y flexibles, y la ubicación especial del mismo, permitió que la versión monomerizada de E2C mantuviese la ‘identidad’ del dímero de referencia (**Figura 1A**). Dichas propiedades fueron confirmadas de acuerdo a lo reportado en ensayos espectroscópicos, cristalográficos (cf. PDB ID: 2Q79) y de unión a DNA medido por ITC y fluorescencia (Dellarole 2005; Dellarole *et al.* 2007). Además, se abordó la caracterización de la estabilidad al equilibrio de scE2C mediante el uso de los agentes químicos Urea y GdmCl. scE2C resultó ser más estable que el dímero de referencia en concentraciones proteicas por debajo de 6 mM (Dellarole 2005; Dellarole *et al.* 2007). Además de los agentes químicos mencionados, utilizamos la temperatura como agente desnaturalizante, aunque por tratarse de una reacción irreversible, no resultó útil a la hora del estudio comparativo entre ambas proteínas. Sin embargo, el análisis en profundidad del comportamiento de scE2C sometida al tratamiento térmico nos otorgó la posibilidad del uso de scE2C como herramienta para ahondar en la ruta amiloidea de E2C y de mecanismos de formación de fibras amiloideas en general, como describimos a continuación.

scE2C como herramienta para el estudio de amiloidosis.

Estudiamos la influencia del aumento de temperatura en E2C y scE2C monitoreando la señal de CD a 225 nm (**Figura 1B**). La señal de CD a esa longitud de onda y normalizada en base al número de enlaces peptídicos, es un fiel reportero del contenido en estructura secundaria de ambas proteínas y por lo tanto del estado de plegamiento en solución. El estado nativo de scE2C y E2C corresponde a una señal de CD a 225 nm de entre 5000 y 6000 $\text{deg.cm}^2.\text{dmol}^{-1}$ mientras que el estado desplegado la señal a dicha longitud de onda tiende a cero (Dellarole 2005; Dellarole *et al.* 2007). Al tratar térmicamente las muestras, en primer lugar nos sorprendió que la curva de desnaturalización térmica no corresponde a una transición clásica. Es esperable que al aumentar la temperatura (\rightarrow), la señal de CD a 225 nm se aproxime a cero y no a señales más negativas como observamos a partir de 60°C (**Figura 1B**). Esto indicaría que las dos proteínas en lugar de perder estructura secundaria o hacerse insolubles disminuyendo la señal, estarían ganando estructura secundaria al ser sometidas al calor. En segundo lugar, observamos que este comportamiento anómalo varía notablemente entre ambas proteínas. Para el caso de E2C (**Figura 1B**, línea gris \rightarrow), el incremento en señal negativa de CD luego de los 60°C disminuye pasados los 70°C mientras que para scE2C no (**Figura 1B**, línea negra \rightarrow). Además al enfriar E2C (**Figura 1B**, línea gris \leftarrow) la señal de CD continúa disminuyendo mientras que al enfriar scE2C (**Figura 1B**, línea negra \leftarrow) la señal de CD se mantiene constante en el valor alcanzado luego de los 60°C. Este patrón de señal de CD se aprecia aún realizando el experimento en un rango de velocidades que van de 1 a 5 $^{\circ}\text{C}/\text{min}$.

Este comportamiento lo observamos también al mantener la temperatura constante en 65°C y medir la señal de CD a 225 nm en función del tiempo (**Figura 1C**). A medida que transcurre el tiempo, la señal negativa de CD de E2C (**Figura 1C**, línea gris) y de scE2C (**Figura 1C**, línea negra \circ) aumenta pero luego de los 1000 segundos, la señal de E2C tiende hacia valores nulos mientras que la señal de scE2C se mantiene negativa. Finalmente, medimos la cinética para scE2C en presencia del sitio de unión de ADN BS2 en una relación molar 1:1 (**Figura 1C**, línea negra \square). Observamos que formando scE2C parte de un complejo con ADN específico, la señal de CD no varía en función del tiempo. Este resultado es acorde a lo observado anteriormente para E2C que es la versión natural dimérica de scE2C (Wetzler *et al.* 2007).

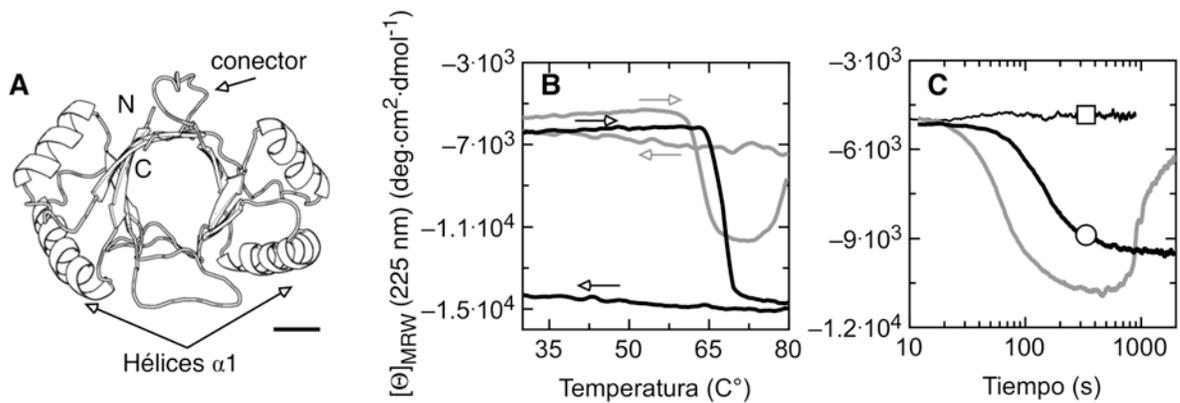


Figura I.1. Tratamiento térmico de E2C y scE2C monitoreado por CD. (A) modelo de la versión monomerizada de E2C, scE2C (Dellarole *et al.* 2007). Barra de escala horizontal: 1 nm. (B) Señal de CD a 225 nm al calentar (→) y enfriar (←) E2C (línea gris) y scE2C (línea negra) a 3°C/min. (C) Cinéticas isotérmicas a 65 °C monitoreadas por la señal de CD a 225 nm de E2C (línea gris) y scE2C (línea negra) en ausencia (○) o presencia (□) de ADN específico (BS2) en una relación 1 a 1. Concentración proteica: 10 μM, tampón químico: 10 mM Hepes pH 7.4, 50 mM NaCl, 1 mM MgCl₂, 1 mM DTT.

Tanto al medir la señal de CD a 225 nm en función de la temperatura como al hacerlo en función del tiempo a temperatura fija, podemos apreciar que los cambios descritos se manifiestan antes en E2C que en la versión monomerizada (**Figura 1B y 1C**). Esto concuerda con lo esperado considerando la estabilidad relativa entre ambas proteínas medidas anteriormente (Dellarole 2005; Dellarole *et al.* 2007). Cabe destacar que al realizar los experimentos de tratamiento térmico, pudimos notar que la pérdida de señal de CD observada para E2C a altas temperaturas o a tiempos largos está acoplada a la formación de agregados visibles. Este comportamiento corresponde por lo tanto a la formación de agregados desordenados irreversibles en el caso de E2C. Por el contrario, en la versión monomerizada no observamos dichos agregados insolubles. Al centrifugar intensivamente la muestra de scE2C tratada térmicamente, la magnitud de la señal de CD no disminuye, indicando la ausencia de agregados insolubles. Esta observación sugeriría que mientras que para el dímero de referencia, E2C, el tratamiento térmico induce la formación de estados agregados insolubles y solubles en simultáneo, para la versión monomerizada, scE2C, se estabiliza únicamente un estado de agregación soluble. Es decir, scE2C estaría ‘congelando’ un estado inestable para la versión dimérica.

El tratamiento térmico induce en scE2C un estado de tipo amiloideo.

A continuación estudiamos si el producto del tratamiento térmico en scE2C presenta

características amiloideas (**Figura 2**). Para ello comparamos espectralmente muestras de scE2C sometidas a tratamiento térmico y posteriormente enfriadas (**Figura 2A**, punto verde), con muestras de scE2C sin tratamiento térmico (**Figura 2A**, punto azul), frente a distintos reporteros de presencia de arquitectura tipo amiloidea: señal de CD, Tioflavina T, Rojo Congo y ANS. Como esperábamos de acuerdo al seguimiento cinético del tratamiento térmico por señal de CD a 225 nm (**Figura 2A**), el espectro completo de CD en el UV cercano de la muestra de scE2C tratada (**Figura 2B**, verde) ganó considerablemente en señal negativa frente a la muestra no tratada (**Figura 2B**, azul). Además, luego del tratamiento térmico, el perfil del espectro de scE2C se modificó en forma contundente. Los dos mínimos espectrales, situados a 210 nm y 224 nm (**Figura 2B**, azul), característicos del estado nativo de E2C (Mok *et al.* 1996) y de su versión monomerizada (Dellarole *et al.* 2007), indican la presencia del componente de estructura helicoidal en presencia de estructura secundaria de tipo hoja beta y de otras contribuciones. Observamos que luego del tratamiento térmico, los dos mínimos desaparecieron y en su lugar apreciamos un solo mínimo espectral situado a 214-218 nm característico de estructura lámina beta (Gray 1996) (**Figura 2B**), característico de estructura amiloidea.

A continuación, incubamos scE2C con o sin tratamiento térmico con Tioflavina T (**Figura 2C**). La Tioflavina T fue el primer compuesto fluorescente utilizado para diagnóstico histológico de agregados amiloideos (Vassar *et al.* 1959). La Tioflavina T en solución, presenta un incremento de varios órdenes de magnitud en la intensidad de fluorescencia al reconocer agregados amiloideos (LeVine 1999). Por esta razón la Tioflavina T es el reportero más utilizado en el campo de la amiloidosis, y hoy en día se sabe que reconoce específicamente la arquitectura característica de agregados amiloideos (Biancalana *et al.* 2010). En las muestras de scE2C tratadas térmicamente, observamos la aparición de señal de fluorescencia de Tioflavina T (**Figura 2C**, azul), mientras que la intensidad de fluorescencia del reportero incubada con scE2C sin tratamiento térmico es nula (**Figura 2C**, verde). Por lo tanto, concluimos que los agregados solubles de scE2C presentan estructura tipo amiloidea.

Luego, incubamos scE2C con o sin tratamiento térmico con Rojo Congo. El Rojo Congo es un colorante industrial que gracias su enigmática propiedad birrefringente de color verde manzana, ganó relevancia a la hora de diagnosticar histológicamente la amiloidosis. En solución, su afinidad por agregados de tipo amiloide se ve acompañada de un corrimiento hacia el rojo en su espectro de absorbancia (Klunk *et al.* 1999). Existen evidencias que muestran que el Rojo

Congo reconoce agregados amiloides en forma similar a detergentes aniónicos (Lendel *et al.* 2010), siendo este correlato de alto interés dado el vínculo entre toxicidad amiloidea y su interacción con membranas biológicas (Luheshi *et al.* 2007). Al incubar con Rojo Congo la muestra de scE2C sometida al tratamiento térmico (**Figura 2D**, verde), observamos un corrimiento en el espectro de absorbancia del Rojo Congo hacia longitudes de onda mayores cuando se lo compara con el espectro del colorante incubado con en la muestra sin tratar térmicamente (**Figura 2D**, azul). Por lo tanto y reforzando el resultado observado con Tioflavina T y CD, los agregados solubles de scE2C presentan estructura tipo amiloidea.

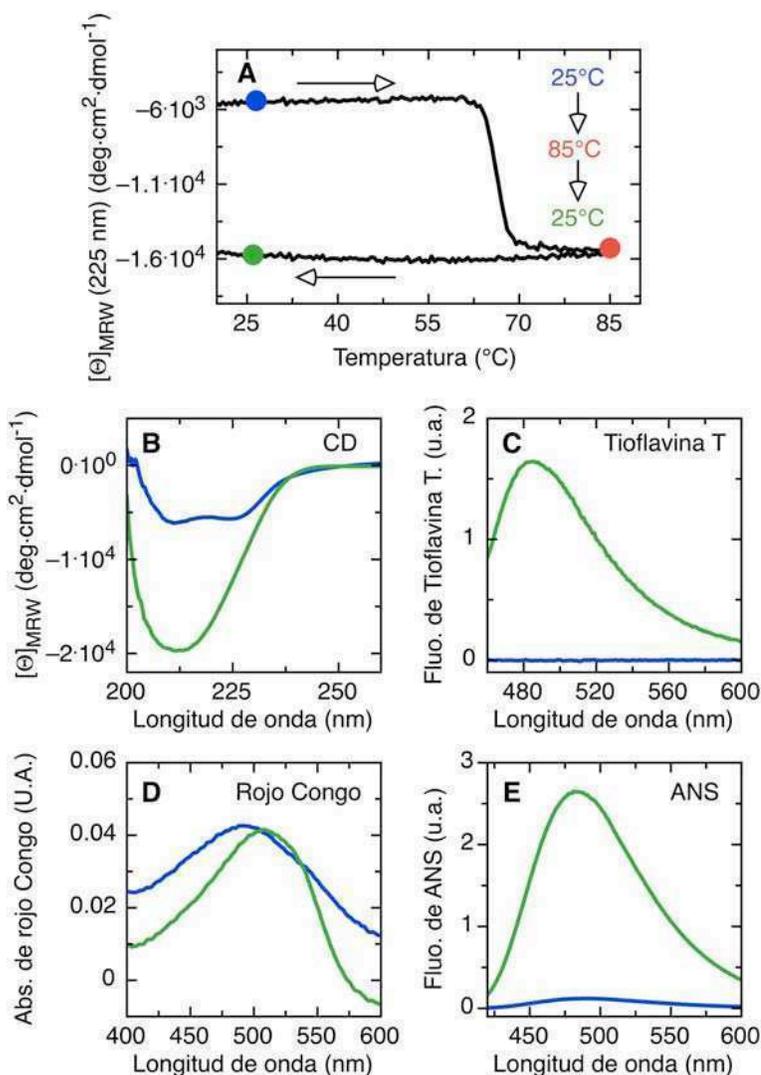


Figura I.2. Características espectroscópicas en scE2C con y sin tratamiento térmico. (A) Señal de CD a 225 nm al calentar (\rightarrow) y enfriar (\leftarrow) scE2C a una velocidad de calentamiento de $3^{\circ}\text{C}/\text{min}$. (B) Espectro de CD. (C) Espectro de fluorescencia de Tioflavina T. (D) Espectro de absorbancia de Rojo Congo. (E) Espectro de fluorescencia de ANS. Coloreado en azul, mediciones realizadas a 25°C sin calentar y en verde, mediciones realizadas a 25°C de una muestra sometida a 85°C . Concentración de scE2C: $10\ \mu\text{M}$, tampón químico: $10\ \text{mM}$ Hepes pH 7.4, $50\ \text{mM}$ NaCl, $1\ \text{mM}$ MgCl_2 , $1\ \text{mM}$ DTT.

Finalmente, medimos la diferencia espectral reportada por la unión de scE2C sometida o no al tratamiento térmico, utilizando el reportero ANS. El ANS es un marcador fluorescente proteico afín a zonas hidrofóbicas accesibles al solvente aunque estructuradas. Por esta propiedad, el ANS se utiliza mayoritariamente para caracterizar estados proteicos parcialmente desplegados, en particular los estados llamados ‘glóbulo fundido’ (Semisotnov *et al.* 1991). El ANS reporta entonces el estado relativo de ‘humedad’ interna proteica (Jha *et al.* 2009). El modo de acción del ANS en cuanto al reconocimiento de agregados amiloideos no se encuentra del todo esclarecido, en principio actuaría como la Tioflavina T dado que ambos compuestos compiten entre sí al encontrarse frente a agregados amiloideos (Biancalana *et al.* 2009). Sin embargo se reportó recientemente que el ANS y la Tioflavina T reaccionan en diferentes estados de polimerización proteica (Bolognesi *et al.* 2010). En efecto, el aumento en la intensidad de la fluorescencia de ANS se encuentra asociado principalmente a una mayor toxicidad del agregado proteico (Bolognesi *et al.* 2010). Los estados de agregación de mayor toxicidad corresponden a estados intermediarios en la formación de fibras amiloideas rígidas, caracterizados por presentar una morfología ‘laxa’, llamados oligómeros o protofibras (Kodali *et al.* 2007). Podemos distinguir en la muestra de scE2C tratada térmicamente e incubada con ANS (**Figura 2E**), un significativo incremento en la intensidad de señal de fluorescencia de ANS (**Figura 2E**, verde), relativo a la intensidad de fluorescencia de scE2C sin tratamiento térmico (**Figura 2E**, azul). Por lo tanto, los agregados solubles de scE2C unen ANS, sugiriendo que presentan una naturaleza amiloidea protofibrilar.

Los cuatro espectros analizados en la **Figura 2** nos indican en conjunto que el tratamiento térmico altera drásticamente la estructura secundaria de scE2C, pasando de estructura nativa a arquitectura de tipo amiloidea, es decir del tipo láminas de hoja beta-cruzada, de morfología tipo protofibrilar. A continuación estudiaremos el tamaño y la morfología que poseen los agregados solubles de scE2C inducidos por tratamiento térmico, mediante el empleo de las técnicas de DLS y AFM.

Diámetro hidrodinámico de los agregados amiloides de scE2C.

Con el fin de analizar el diámetro hidrodinámico de las especies solubles de scE2C formadas por el efecto de la temperatura, medimos la dispersión de la luz en función del tiempo o dispersión dinámica (DLS, acrónimo de sus siglas en inglés) de scE2C sometida a tratamiento

térmico (**Figura 3**). Las variaciones en intensidad de la luz dispersada por moléculas en solución permite estimar el diámetro de las mismas. Brevemente, el ruido de la luz dispersada reporta la velocidad del movimiento Browniano de las partículas presentes y mediante la relación Einstein-Stokes podemos calcular el diámetro. La relación de Einstein-Stokes explica la difusión de partículas esféricas a través de líquidos, por lo tanto todas las especies medidas en DLS son consideradas esféricas. Del mismo modo que estudiamos la señal de CD a 225 nm en función del tiempo para el tratamiento térmico de scE2C a velocidad de calentamiento de 3°C/min, en el rango de temperaturas de 25°C a 85°C (**Figuras 2A y 3A**), medimos el radio hidrodinámico de scE2C (**Figura 3B**), con el fin de correlacionar el cambio en la estructura secundaria de scE2C observado por CD con cambios en el diámetro hidrodinámico de scE2C. En la **Figura 3B**, observamos que se mantiene estable un diámetro hidrodinámico de 4.1 ± 1.2 nm en el rango de temperaturas de 25°C a 57°C (barras azules). Este resultado concuerda con una señal de CD a 225 nm característica de scE2C en estado nativo (**Figura 3A**, línea azul) y con el diámetro de la molécula medido en la estructura cristalográfica de scE2C (PDB: 2Q79 (Dellarole *et al.* 2007)), de ~ 4 nm. (**Figura 1A**). Sin embargo, al superar los 57°C de temperatura, el pico correspondiente a ~ 4 nm desaparece y observamos dos picos dispersos correspondientes a diámetros hidrodinámicos de 10.3 ± 5.2 nm y 52.1 ± 40 nm (**Figura 3B**, barras rojas). Por un lado, la temperatura a la cual aparecen especies de gran diámetro hidrodinámico coincide con la temperatura de caída en señal de CD (**Figura 3A**, línea roja). Esta correlación sugiere que las especies con características amiloideas son oligoméricas. Por otro lado, la amplitud de los picos que apreciamos en el rango de temperaturas de 57°C a 85°C insinúa la presencia de una gran variedad de especies.

Para refinar la cuantificación del diámetro hidrodinámico de las especies formadas y confirmar la solubilidad de las mismas, medimos DLS de la proteína scE2C a 25°C luego de ser llevada a 85°C y centrifugada intensivamente (**Figura 3C y Tabla 1**). Por un lado, observamos que el pico medido a 25°C sin calentar correspondiente a un diámetro hidrodinámico de scE2C de 4.2 ± 1.2 nm (**Figura 3C**, barras azules) no se recupera luego de calentar a 85°C y enfriar la muestra a 25°C (**Figura 3C**, barras verdes). Por otro lado, al enfriar la muestra de scE2C observamos tres picos: ① de diámetro hidrodinámico de 10 ± 1 nm, ② de diámetro hidrodinámico 26 ± 8 nm y ③ de diámetro hidrodinámico 59 ± 1 nm (**Figura 3C**, barras verdes). Los picos de la **Figura 3C** se corresponden con los de la **Figura 3B** pero con mayor definición.

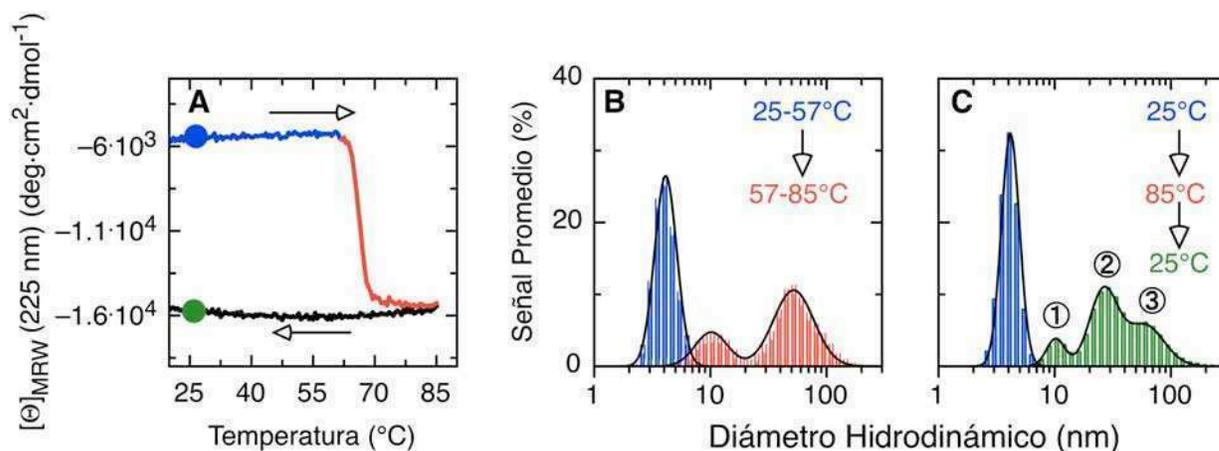


Figura I.3. Caracterización del diámetro hidrodinámico de scE2C con y sin tratamiento térmico. (A) Señal de CD a 225 nm al calentar (\rightarrow) y enfriar (\leftarrow) scE2C a una velocidad de calentamiento de 3°C/min. En azul señal de CD en el rango 25°C a 57°C y en rojo de 57°C a 85°C. (B) Tratamiento térmico de scE2C seguido por DLS a una velocidad de calentamiento de 3°C/min. Barras azules: diámetro promedio medido en el rango de temperaturas de 25°C a 57°C, valor del pico: 4.1±1.6 nm. Barras rojas: diámetro promedio medido en el rango de temperaturas de 57°C a 85°C, valor de los picos: 10.2±5.2 nm y 52±40 nm. (C) Diámetro de scE2C con (verde) y sin (azul) tratamiento térmico medido a 25°C (Tabla 1). En DLS, las líneas negras corresponden al ajuste de los datos a una función gaussiana, concentración de scE2C: 16.5 μM, tampón químico: 10 mM Hepes pH 7.4, 50 mM NaCl, 1 mM MgCl₂, 1 mM DTT.

Tabla I.1. Diámetro hidrodinámico de scE2C sometida al tratamiento térmico.

Especies	Diámetro hidrodinámico* (nm)	$Ve^{\$}$ (nm ³)	$Ve^{Pico} / Ve^{scE2C \dagger}$ (nm ³)
scE2C	4.2±1.2	39±12	1
Pico ①	10.2±3.8	555±67	14±5
Pico ②	26.3±8.6	9525±390	246±23
Pico ③	59.3±25.2	105200±2600	2814±255

* Los picos corresponden a los representados en la Figura 3C. El diámetro hidrodinámico fue calculado en base al ajuste de los datos de DLS a una gaussiana, el error corresponde a dos veces el desvío estándar, intervalo de confianza 95%. $Ve^{\$}$: volumen de una esfera: $Ve = 4/3 \cdot \pi \cdot r^3$, siendo r el radio hidrodinámico medido, el error proviene de la correspondiente propagación. \dagger Número de scE2C presentes en cada Pico asumiendo que cada Pico corresponde a una morfología esférica de volumen: Ve^{Pico} , el error proviene de la correspondiente propagación.

Dado que la técnica de DLS interpreta a las especies como esferas, calculamos en base al volumen correspondiente el número de moléculas de scE2C que conformarían cada especie (Tabla 1). Indudablemente de no tratarse realmente de esferas estaríamos sobre o subestimando dicha estimación. En una esfera de 60 nm de diámetro como sería la especie ③, considerando al monómero como una entidad flexible o líquida, entrarían unas 3000 moléculas de scE2C (Tabla

1). Dicha entidad tendría un peso molecular de 60 megadaltons, un valor dudoso para una esfera proteica soluble. Como referencia, el peso molecular del complejo proteico del poro nuclear es de entre 40 y 60 MDa y de diámetro de 90 a 120 nm y altura de 30 a 50 nm (Brohawn *et al.* 2009). Por esta razón para comprender mejor los resultados de DLS es necesario definir las morfologías presentes en la muestra de agregados solubles de scE2C, para ello, a continuación utilizamos la técnica de microscopía de fuerza atómica (AFM, acrónimo de sus siglas en inglés).

Caracterización morfológica de los agregados amiloides de scE2C.

Analizamos las especies formadas al tratar térmicamente scE2C en escala nanométrica mediante la técnica de microscopía de barrido de microscopía de fuerza atómica (**Figura 4**). La técnica de AFM consiste en el escaneado de la muestra en tres dimensiones mediante un pico o *tip* de punta muy fina. Primero, medimos scE2C sin tratar y verificamos la pureza del tampón químico de medición (**Figura 4A**). Como esperábamos, scE2C sin tratar y a una concentración de 20 ng/μl (equivalente a una concentración de 1 μM), forma una capa proteica homogénea sobre el soporte de mica, imperceptible para la resolución del AFM (**Figura 4A**). Al medir la muestra de scE2C sometida al tratamiento térmico, llevada a 85°C y luego enfriada, observamos la presencia de especies aisladas con forma de ‘tipo-gusanos’ junto a la presencia de especies de gran tamaño constituidas de aglomerados de gusanos aislados (**Figura 4B**). El soporte de mica no es inocuo frente a procesos de polimerización de tipo amiloideo (Hoyer *et al.* 2004). Para analizar si los aglomerados corresponden a una morfología presente en solución o si corresponden a un aglomerado inducido por el soporte y la alta concentración de sembrado, redujimos la concentración de scE2C sembrada. Al disminuir la cantidad sembrada de los agregados solubles de scE2C de 20 ng/μl a 2 ng/μl (**Figura 4C**) pasando por 10 ng/μl (**Figura 4B**, recuadro), notamos que la cantidad de aglomerados de gran tamaño disminuye hasta desaparecer, mientras que los agregados de tipo gusano aislado continúan presentes. Esto corrobora que las especies de gran tamaño no se corresponden con una morfología específica sino con aglomerados de gusanos formados en la superficie del soporte de medición.

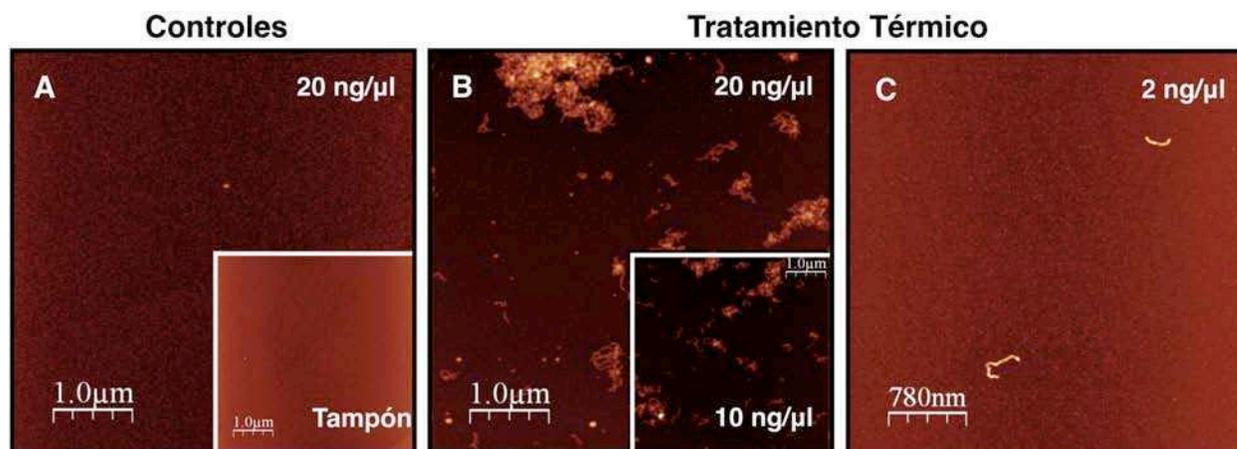


Figura I.4. Microscopía de Fuerza Atómica de scE2C con y sin tratamiento térmico. (A) Campo de 5x5 μm de scE2C sin tratamiento térmico a una concentración de 20 $\text{ng}/\mu\text{l}$, recuadro, campo de 5x5 μm del tampón químico (B) Campo de 5x5 μm de scE2C luego de tratamiento térmico a una concentración de 20 $\text{ng}/\mu\text{l}$ y 10 $\text{ng}/\mu\text{l}$, recuadro. (C) Campo de 3.9x3.9 μm de scE2C luego de tratamiento térmico a una concentración de 2 $\text{ng}/\mu\text{l}$. Las muestras fueron medidas en el equipo de AFM del Centro de Microscopías Avanzadas, FCEN, UBA, y analizadas mediante el operador lógico WSXM (Horcas *et al.* 2007). Todas las muestras fueron filtradas, centrifugadas y fijadas en un soporte de mica en el tampón químico: 10 mM Hepes pH 7.4, 50 mM NaCl, 1 mM MgCl_2 .

Para definir que morfologías predominan, ampliamos la resolución de las mediciones de AFM bajando el tamaño del campo escaneado y midiendo las muestras de los agregados solubles de scE2C en condiciones húmedas y en condiciones secas (**Figura 5**, ver métodos). Identificamos tres tipos de morfología preponderantes: tipo bola pequeña (**Figura 5A**, ①), tipo bola mayor (**Figura 5B**, ②) y tipo gusano (**Figura 5C**, ③). Estos últimos, están compuestos por unidades de forma de bola distribuidas con una periodicidad de unos 40 nm (**Figura 5D**). El ancho de las bolas mayores, de 39 ± 7 nm, coincide con el ancho de las unidades formadoras de los gusanos, pero no el ancho de las bolas pequeñas, 15.5 ± 2.6 nm (**Tabla 2**). El tamaño de los gusanos resultó ser heterogéneo pudiéndose distinguir gusanos desde 100 a 600 nm de longitud, es decir compuestos por 2 a 15 bolas mayores o subunidades.

Además del ancho y longitud de contorno analizamos la altura (eje z) de las tres morfologías descriptas con el fin de estimar su volumen. A diferencia de la técnica de DLS, la fuerte interacción de las especies con el soporte de medición en AFM deforma las geometrías de las mismas aplanándolas. Por ende al estimar volúmenes consideramos a las bolas como cilindros o ‘monedas’ y a los gusanos como cilindroides o cilindros elípticos (**Tabla 2**). La morfología tipo bola pequeña (**Figura 5A**, ①), corresponde a una esfera de volumen aproximado de 558 nm^3 y en la que entrarían unas 12 moléculas de scE2C (**Tabla 2**). Este volumen concuerda excelentemente con el pico ① medido en DLS (**Tabla 1**). A su vez, el volumen de la

morfología tipo bola mayor estimado en AFM de aproximadamente 7000 nm^3 (**Figura 5B** ② y **Tabla 2**) concuerda dentro del error con el volumen del pico ② medido en DLS (**Tabla 1**). Por lo tanto, la morfología de tipo bola mayor podría incluir unas 200 moléculas de scE2C o 15 bolas pequeñas. Sin embargo, aún considerando el gusano de mayor longitud de contorno el volumen estimado en AFM (**Figura 5C** ③ y **Tabla 2**), no concuerda con el pico ③ medido en DLS (**Tabla 1**), en buen acuerdo con una sobreestimación técnica al considerar esferas.

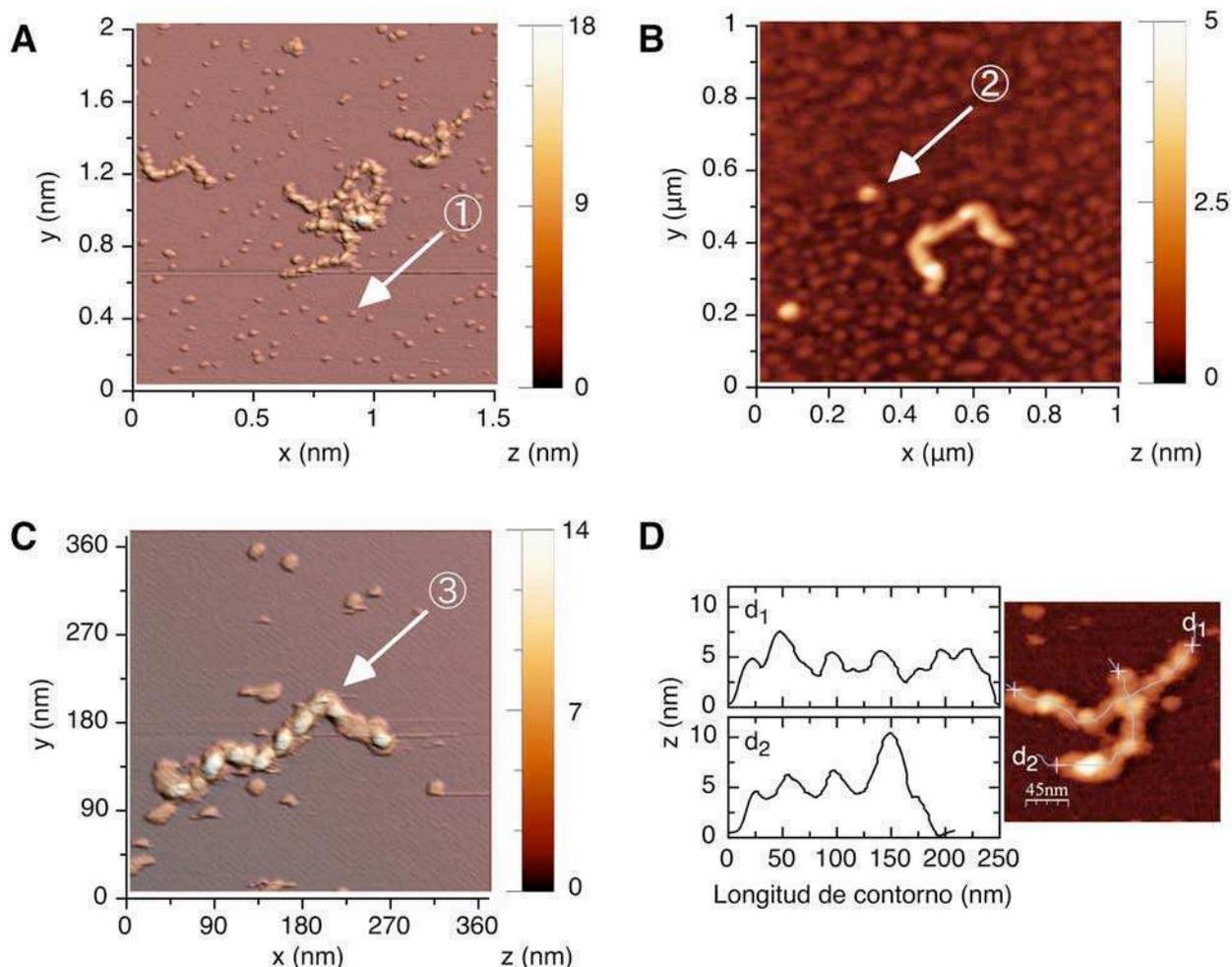


Figura I.5. Caracterización morfológica de los agregados solubles de scE2C por AFM. (A) Campo de $1.5 \times 1.5 \mu\text{m}$ medido en solución, la flecha indica un ejemplo de agregados de morfología tipo bola pequeña (①). (B) Campo de $1 \times 1 \mu\text{m}$ medido en seco, la flecha indica un ejemplo de agregados de morfología tipo bola mayor (②). (C) Campo de $0.36 \times 0.36 \mu\text{m}$ medido en solución, la flecha indica un ejemplo de agregados de morfología tipo gusano (③). (D) Ilustración del trazado de la longitud de contorno de dos agregados tipo gusano superpuestos, d_1 y d_2 . Todas las muestras conteniendo una concentración de scE2C de $10 \text{ ng}/\mu\text{l}$, fueron filtradas, centrifugadas y fijadas en un soporte de mica en el tampón químico: 10 mM Hepes pH 7.4, 50 mM NaCl, 1 mM MgCl_2 .

Cabe mencionar, por un lado, que se pudieron observar como describimos en el esquema final, algunas duplas y ternas de bolas pequeñas, indicando que las bolas pequeñas podrían

interactuar entre sí para asociarse y formar las bolas mayores. Por otro lado, observamos que el ancho de las bolas mayores (**Tabla 2**) coincide notablemente con la distancia entre picos observados en los gusanos (**Figura 5C**). Las diferencias en la cantidad de bolas pequeñas y mayores observadas (ver leyenda de **Tabla 2**), indicarían que las bolas mayores corresponderían a una morfología de menor tiempo de vida o estabilidad que las pequeñas, probablemente porque se ensamblarían rápidamente para formar los gusanos de tipo ‘collar de perlas’. Podemos confirmar combinando lo descrito hasta aquí, que las tres morfologías presentan características amiloideas y se encuentran en solución dada la correspondencia entre DLS y AFM y a su vez poseen cambios espectrales característicos de agregados amiloideos. A continuación, estudiaremos la raíz que da origen a la formación del paisaje morfológico de especies producto de la polimerización de scE2C inducida por calor.

Tabla I.2. Caracterización de las especies morfológicas encontradas en AFM.

Morfología	Ancho (nm)	Altura (nm)	Longitud de contorno (nm)	V_c^{AFM} (nm ³)	$\frac{V_{e^{scE2C}}}{V_c^{AFM}}$ †
① Bola Pequeña*	15.5±2.6	2.9±0.3	-	558±144	12±6
② Bola Mayor*	39±7	4.7±2	-	6800±3800	181±102
③ Tipo Gusano [#]	30±7	3.2 ± 0.3	95 a 576	14000 a 86000	380 a 2300

La altura corresponde al máximo de intensidad y el ancho a la distancia al 50% del máximo de intensidad (ancho total máximo medio, FWHM) y la longitud de contorno mediante el operador lógico WSXM (Horcas *et al.* 2007). $N_{bola\ chica} = 70$; $N_{bola\ mayor} = 30$; $N_{gusanos} = 12$. Los datos de las bolas chicas presentaron una distribución normal y el valor corresponde a la media aritmética proveniente del ajuste a una gaussiana, siendo el error el correspondiente desvío estándar. Los datos de bolas mayores y de los gusanos corresponden al valor promedio y su correspondiente desvío estándar. *Las bolas fueron consideradas como cilindros de volumen $V_c = \pi \cdot r^2 \cdot h$, siendo r el radio (ancho/2) y h la altura, el error proviene de la correspondiente propagación. [#]Los gusanos fueron considerados como cilindroides o cilindros elípticos de volumen $V_{ce} = \pi \cdot r \cdot h \cdot cl$, siendo r el radio tomado de el ancho/2, h la altura y cl la longitud de contorno, el error proviene de la correspondiente propagación. † Número de scE2C presentes en cada morfología calculada a partir del volumen esférico V_e de scE2C y el volumen cilíndrico V_c estimado en cada morfología, el error proviene de la correspondiente propagación.

Mecanismo de polimerización de scE2C.

Los compuestos químicos reporteros de agregaciones ordenadas pueden interferir en el evento de polimerización, complicando el estudio del mecanismo (Lorenzo *et al.* 1994). Resulta conveniente monitorear la reacción de polimerización con la menor intervención posible

(O'Nuallain *et al.* 2006). Es por ello que de las señales analizadas en la **Figura 2**, decidimos medir el cambio en la señal de CD como señal reportera a la hora de realizar un análisis cinético detallado del agregado amiloideo de scE2C. La señal de CD es muy sensible a interferencias por dispersión de la luz debido a agregados de gran tamaño (Colon 1999), sin embargo como reportamos recién, los agregados amiloideos de scE2C son solubles, por lo que presentan baja dispersión.

Medimos la señal de CD a 225 nm en función del tiempo en un rango de concentración de scE2C de 1 a 25 μM (**Figura 6**). Las cinéticas isotérmicas fueron medidas a 65°C (**Figura 6A**) y 70°C (**Figura 6B**) (ver diseño experimental en Materiales y Métodos). Podemos describir las cinéticas mediante tres fases subsecuentes: una fase corta sin cambios en la señal de CD (*fase de retraso o nucleación*) seguida de una fase larga donde se registran cambios en señal de CD (*fase de crecimiento o elongación*) que termina al estabilizarse la señal de CD (*fase de equilibrado o terminación*). Estas características son compatibles con un mecanismo de polimerización nucleada en el cual la formación de un núcleo o semilla es energéticamente desfavorable, *fase de retraso*, y es seguido por una etapa de polimerización que es energéticamente favorable, *fase de elongación* (Ferrone 1999). Cualitativamente, observamos que al aumentar la concentración de scE2C a 65°C (**Figura 6A**), la caída de señal negativa de CD comienza a tiempos menores y la pendiente de la transición se vuelve más empinada. En las isotermas medidas a 70°C (**Figura 6B**), observamos que no se conforman dichos cambios al variar la concentración. Por otra parte observamos que los tiempos de finalización de las curvas es de uno a dos órdenes menor a 70°C que a 65°C. La formación del núcleo y la velocidad de elongación están dictaminados por la concentración y cambio conformacional de la proteína; a continuación analizaremos las cinéticas en base a dos parámetros característicos de cada etapa con el fin de clarificar el mecanismo de polimerización observado.

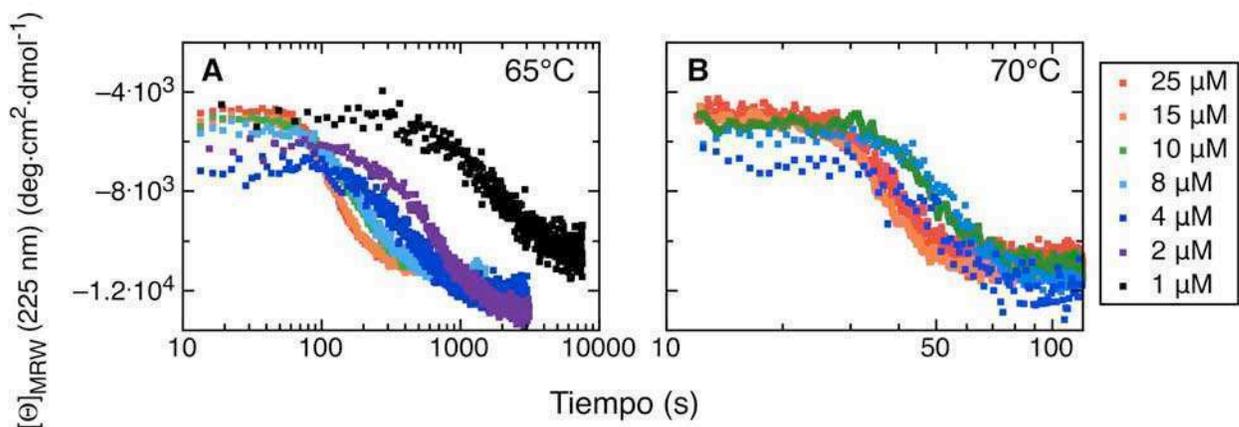


Figura I.6. Cinéticas isotérmicas en función de la concentración.(A) Cinéticas isotérmicas a 65°C monitoreadas por señal de CD a 225 nm de sCE2C en un rango de concentración proteica de 1 a 25 μM . (B) Cinéticas isotérmicas a 70°C monitoreadas por señal de CD a 225 nm de sCE2C en un rango de concentración proteica de 4 a 25 μM . Las distintas concentraciones están coloreadas de acuerdo al panel de la derecha. Tampón químico de medición: 10 mM HEPES pH 7.4, 50 mM NaCl, 1 mM MgCl₂, 1 mM DTT.

El análisis cinético de la nucleación se caracterizó en base al tiempo requerido para que la reacción llegue al 50% del cambio total en la señal de CD: t_{50} . El parámetro t_{50} reporta en forma robusta el tiempo de la *fase de retraso* y por ende el proceso de nucleación que desencadena la elongación. En la **Figura 7A** ilustramos la extracción de una cinética isotérmica del parámetro t_{50} . Los valores de t_{50} medidos en función de la concentración, se encuentran representados en la **Tabla 3**. En una polimerización nucleada clásica, t_{50} muestra una fuerte dependencia en la concentración total de proteína (Goldstein *et al.* 1986; Ferrone 1999), de acuerdo a la siguiente relación log-log:

$$\log_{10} t_{50} = Q - \frac{n+1}{2} \times \log_{10} C_{total} \quad [1]$$

siendo Q una constante, n el número de monómeros en el núcleo y C_{total} la concentración total de proteína. Sin embargo, apreciamos que las cinéticas son independientes de la concentración a 70°C y presentan una dependencia leve y curvada a 65°C (**Figura 7B**). La curvatura en el log-log plot de t_{50} versus C_{total} fue interpretado recientemente (Powers y Powers 2006), atribuyendo el efecto a una estabilización del núcleo oligomérico a altas concentraciones de proteína. Dicha estabilización reduce la proporción de monómero disponible para la reacción de polimerización por lo que la C_{total} en la ecuación [1], se corrige a una concentración efectiva para la polimerización ($C_{efectiva}$) de acuerdo con:

$$C_{efectiva} = \frac{K_s(2C_{total} + K_s - \sqrt{4K_s C_{total} + K_s^2})}{2C_{total}} \quad [2]$$

dónde K_s es la ‘concentración supercrítica’ a partir de la cual el núcleo oligomérico se puebla. La relación log-log de t_{50} versus C_{total} [1] se corrige a:

$$\log_{10} t_{50} = Q - \frac{n+1}{2} \times \log_{10} \left[\frac{K_s(2C_{total} + K_s - \sqrt{4K_s C_{total} + K_s^2})}{2C_{total}} \right] \quad [3]$$

De esta forma, en base al valor de K_s y al de la concentración total del experimento existen tres situaciones posibles: una fuerte dependencia de t_{50} con la concentración si $C_{total} < K_s$ (ver ecuación[1]), una leve dependencia de t_{50} con la concentración si $K_s \approx C_{total}$ y una nula dependencia de t_{50} con la concentración si $C_{total} > K_s$ (Powers y Powers 2006). Utilizando el concepto de ‘concentración supercrítica’, pudimos describir adecuadamente el evento de nucleación de la polimerización de scE2C (**Figura 7B**).

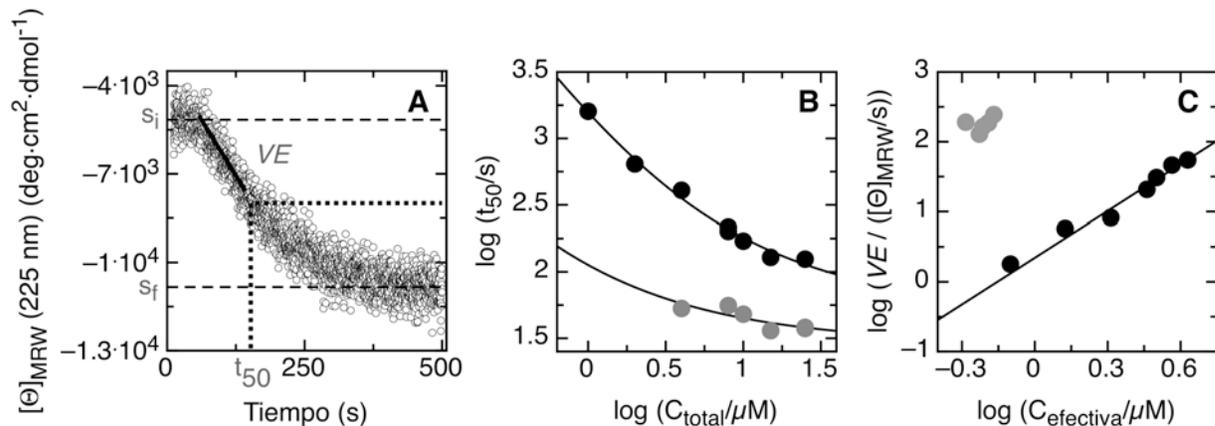


Figura I.7. Evaluación y dependencia cinética de t_{50} y VE con la concentración de scE2C. (A) Ilustración del cálculo de los parámetros extraídos de las cinéticas. s_i y s_f fueron calculados ajustando el conjunto de datos iniciales o finales a una constante (líneas discontinuas). t_{50} : valor de la señal de CD al 50% del cambio total de señal (líneas punteadas). VE fue calculado ajustando el conjunto de valores de la señal de CD al 33% del cambio total de señal a una recta, siendo VE su pendiente (línea continua) (B) Gráfico Log-log mostrando la dependencia de t_{50} con la concentración total de scE2C para las isotermas medidas a 65°C (negro) y 70°C (gris), la línea corresponde al ajuste de los datos a la ecuación [3] correspondiente al modelo de ‘concentración supercrítica’ (Powers y Powers 2006). (C) Gráfico Log-log mostrando la dependencia de VE con la concentración total de scE2C para las isotermas medidas a 65°C (negro) y 70°C (gris), la concentración de scE2C fue corregida a $C_{efectiva}$ de acuerdo a la ecuación [2] correspondiente al modelo de ‘concentración supercrítica’ (Powers y Powers 2006).

El Análisis cinético de la elongación se llevó a cabo estudiando la dependencia de la velocidad de elongación (VE) con la concentración de la proteína scE2C. El parámetro cinético

VE corresponde a la mayor pendiente en el cambio de señal de CD, en la **Figura 7A** ilustramos su cálculo. Los valores de VE medidos en función de la concentración se encuentran representados en la **Tabla 3**. VE nos reporta sobre la *fase de crecimiento* siendo una medida de la velocidad de reducción de monómero scE2C libre, es decir a mayor VE más rápido el monómero scE2C se incorpora en un agregado amiloideo. Si la elongación de un agregado amiloideo preexistente ocurre por la adición de un oligómero scE2C conformado de tamaño l entonces,

$$VE = -k_{elongación} C_{efectiva}^l \quad [4]$$

siendo $C_{efectiva}$ la concentración de scE2C monomerizada disponible, calculada de acuerdo a la ecuación [2] mediante las K_s medidas. Dado que la concentración del polímero final es la misma al comienzo de las reacciones, el gráfico log-log de VE versus $C_{efectiva}$ tendrá l como pendiente. En la **Figura 7C**, graficamos la dependencia de la VE con la $C_{efectiva}$. La velocidad de elongación es independiente de la concentración efectiva a 70°C, como esperábamos de acuerdo al modelo de nucleación utilizado (Powers y Powers 2006). La reacción de elongación a 65°C posee un orden de reacción de 2.2 ± 0.1 , indicando que la elongación ocurre vía la adición de dos monómeros al agregado amiloideo. Esto concordaría con el modelo de Powers y Powers en una situación de $K_s > C_{total}$, es decir presencia de un núcleo estable a mayores concentraciones que la ‘concentración supercrítica’.

Tabla I.3. Parámetros de las cinéticas isotérmicas en dependencia con la concentración.

Concentración de scE2C μM	t_{50} s	VE [Θ] _{MRW} /s	s_i [Θ] _{MRW}	s_f [Θ] _{MRW}
1 ^a	1600±279	-1.8±0.1	-5069±74	-10351±28
2 ^a	630±51	-5.7±0.5	-6196±58	-12729±17
4 ^a	407±66	-8.2±0.4	-6959±52	-12167±10
8 ^a	199±18	-19±1	-4791±29	-9900±8
10 ^a	146±9	-33±2	-5642±92	-12399±34
15 ^a	128±3	-46±2	-5096±27	-11107±9
25 ^a	124±2	-55±3	-4915±20	-10822±7
4 ^b	53±5	-94±21	-6879±80	-12303±26
8 ^b	56±2	-128±13	-5822±28	-11620±26
10 ^b	42±2	-155±59	-5529±44	-11456±16
15 ^b	36±1	-186±13	-5035±50	-11263±21
25 ^b	37±1	-245±31	-4880±25	-10629±10

^a isotermas cinéticas medidas a 65°C, ^b isotermas cinéticas medidas a 70°C. Los parámetros t_{50} , VE , s_i y s_f fueron obtenidos de hasta tres experimentos, ver **Figura 7A** y métodos. [Θ]_{MRW} es una reducción de [Θ]_{MRW(225 nm)} (deg.cm².dmol⁻¹).

La nucleación y la elongación están acopladas.

Por un lado, determinamos que la dependencia de los parámetros t_{50} y VE con la concentración de scE2C no es la misma a 65°C que a 70°C . A continuación, extendemos el estudio de la dependencia de los parámetros t_{50} y VE para el rango de temperaturas de 60°C a 72.5°C a una concentración fija de scE2C de $10\ \mu\text{M}$ (**Figura 8**). La dependencia de la temperatura nos permite establecer la energía de activación del mecanismo de polimerización en estudio de acuerdo a la ecuación de Arrhenius (ver Materiales y Métodos). En primer lugar certificamos que en todas las mediciones, el estado inicial y el final de las isothermas medidas sea la misma como lo reportan la señal de CD a $225\ \text{nm}$ s_i y s_f respectivamente (**Figura 8A** y **Tabla 4**). Tanto el gráfico de pseudo-Arrhenius entre t_{50} y T^{-1} (**Figura 8B**) como el gráfico de pseudo-Arrhenius entre VE y T^{-1} (**Figura 8C**), muestran un cambio en la pendiente alrededor de los 67°C . La mayor pendiente corresponde a una mayor energía de activación de la reacción, y se ubica en el rango de temperaturas más bajas, incluyendo la cinética medida a 65°C (**Figura 7B** y **7C** en negro). Mientras tanto, la menor energía de activación abarca las temperaturas más altas, incluyendo la cinética medida a 70°C (**Figura 7B** y **7C**, en gris).

En el rango de temperaturas analizadas, dicho cambio en la pendiente sugiere un cambio en la activación de la reacción de polimerización, coincidente con el estudio de la dependencia de la concentración donde observamos que a 65°C (**Figura 7B** y **7C** en negro) el núcleo es menos estable que a 70°C (**Figura 7B** y **7C** en gris). El cambio en energía de activación es un reflejo del cambio en la ‘concentración supercrítica’ observado en el estudio de las cinéticas de polimerización en dependencia de la concentración. El encontrarse muy por encima de la concentración supercrítica a 70°C , coincide con la presencia de una baja energía de activación, mientras que encontrarse a concentraciones cercanas a la concentración supercrítica coincide con la presencia de una alta energía de activación a 65°C . Finalmente, en la **Figura 8D** observamos una estricta correlación entre todos los parámetros t_{50} y VE calculados en el presente estudio cinético (**Tablas 3** y **4**) indicando que en el proceso de polimerización de scE2C inducido por temperatura, la elongación y la nucleación son dos mecanismos acoplados (Fandrich 2007). La presencia de una absoluta correlación entre la nucleación y la elongación reafirma también el modelo de nucleación implementado, considerando el núcleo estructural conformado por 2 ± 1 monómeros aún en presencia de cambios en la concentración supercrítica.

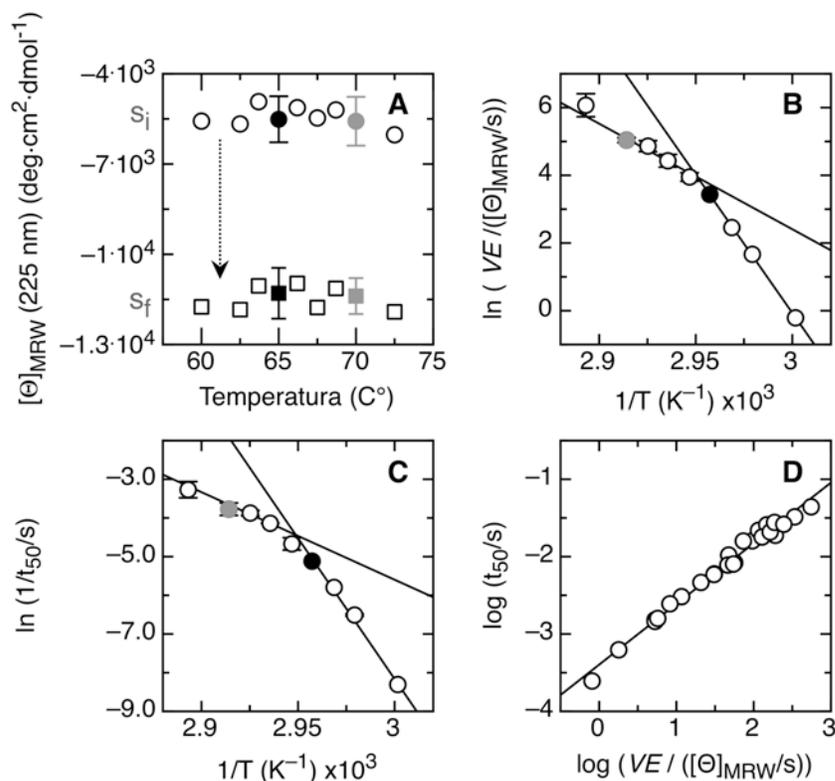


Figura I.8. Dependencia cinética de t_{50} y VE con la temperatura. (A) Valor de las señales iniciales, s_i , (●) y finales, s_f , (■) de CD a 225 nm de cinéticas isotérmicas medidas entre 60°C y 72.5°C a una concentración de scE2C de 10 μ M (círculos y cuadrados vacíos), ver además **Tabla 4**. Para las cinéticas medidas a 65°C y 70°C a una concentración de scE2C de 1 a 25 μ M graficamos el valor promedio y su desvío estándar (círculos y cuadrados rellenos en gris y negro respectivamente), ver además la **Tabla 3**. (B) Gráfico de pseudo-Arrhenius para la velocidad de elongación, VE . Las dos rectas representan el ajuste a un polinomio de grado uno. (C) Gráfico de pseudo-Arrhenius para el tiempo al 50% de caída en señal de CD, t_{50} . Las dos rectas corresponden a un ajuste a un polinomio de grado uno. (D) Gráfico log-log de t_{50} vs. VE , la recta consiste en el ajuste a un polinomio de grado uno.

Tabla I.4. Parámetros de las cinéticas isotérmicas en dependencia con la temperatura.

Temperatura	t_{50}	VE	s_i	s_f
C°	s	$[\Theta]_{MRW}/s$	$[\Theta]_{MRW}$	$[\Theta]_{MRW}$
60	4038±298	-0.8±0.1	-5574±48	-11747±23
62.5	660±16	-5.3±0.6	-5574±48	-11858±15
63.7	317±21	-12±1	-4927±37	-11049±11
65	146±6	-33±2	-5642±91	-12399±34
66.2	105±4	-52±5	-5127±43	-10967±15
67.5	61±1	-85±21	-5351±41	-11545±15
68.7	49±2	-130±17	-5348±47	-11825±27
70	42±2	-156±59	-5348±39	-11390±28
72.5	26±1	-447±79	-5202±33	-11133±10

Todas las cinéticas fueron realizadas a una concentración de 10 μ M. Los parámetros t_{50} , VE , s_i y s_f fueron calculados en base a como máximo tres experimentos. $[\Theta]_{MRW}$ es una reducción de $[\Theta]_{MRW(225\text{ nm})}$ (deg.cm².dmol⁻¹).

Caracterización morfológica por microscopía electrónica.

El análisis cinético sistemático realizado indicó la existencia de un núcleo de 2 ± 1 monómeros de scE2C. De acuerdo al modelo cinético, dicho núcleo sería lo suficientemente estable y estructurado como para mantenerse presente a lo largo del evento de polimerización. Frente a este panorama, decidimos estudiar morfológicamente las muestras de agregados solubles de scE2C mediante microscopía electrónica con el fin de reanalizar el paisaje morfológico en búsqueda de especies más pequeñas que las observadas por AFM (**Figura 9**). Para ello, enviamos muestras de scE2C con y sin tratamiento térmico al laboratorio de Estructura y Función Molecular (Colegio Imperial de Londres) en colaboración con la especialista en criomicroscopía Dra Cecilia Bebeaqua. Medimos las muestras por criomicroscopía (**Figura 9A**) y por tinción negativa (**Figura 9B**) (NS, acrónimo de sus siglas en inglés). En ambos métodos de visualización por microscopía electrónica, se distinguen morfologías de tipo bola (flechas blancas) y de tipo gusano (flechas negras) del mismo tamaño a las observadas por AFM.

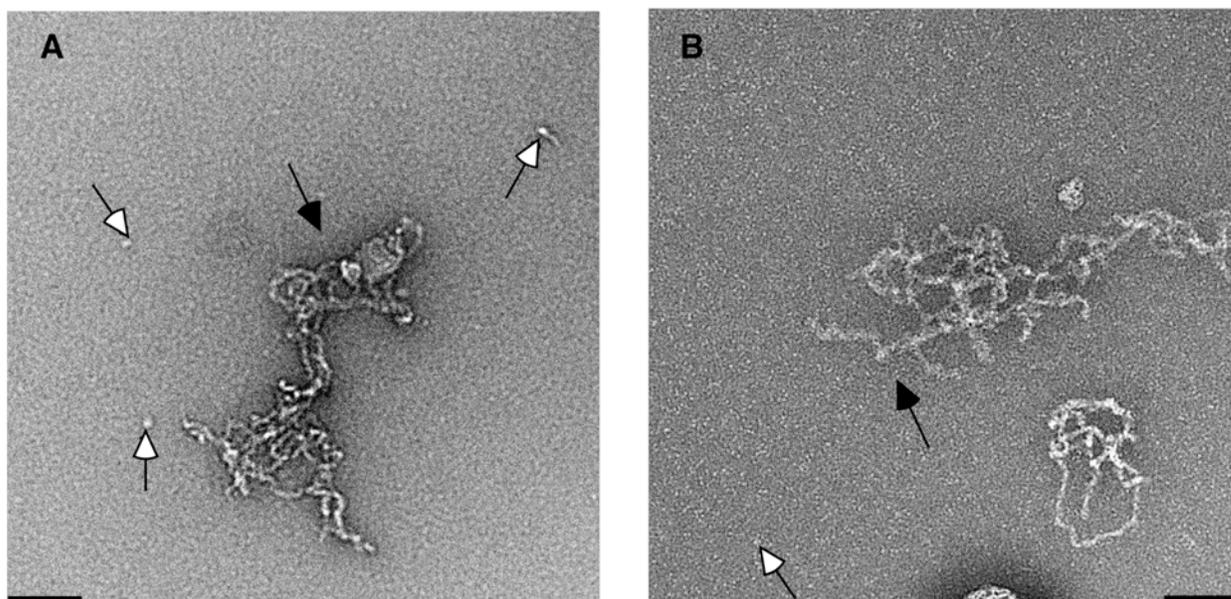


Figura I.9. Análisis morfológico de los agregados solubles de scE2C por microscopía electrónica. (A) Campo de $0.73 \times 0.73 \mu\text{m}$ de scE2C con tratamiento térmico medido por criomicroscopía. (B) Campo de $0.73 \times 0.73 \mu\text{m}$ de scE2C con tratamiento térmico medido por tinción negativa. Las flechas blancas y negras señalan morfologías de tipo bola y gusano, respectivamente. Barra de escala horizontal: 90 nm. Todas las muestras conteniendo una concentración de scE2C de 0.5 mg/ml, fueron filtradas, centrifugadas y medidas en el tampón químico: 10 mM Hepes pH 7.4, 50 mM NaCl, 1 mM MgCl_2 .

A diferencia de la microscopía electrónica por tinción negativa cuya resolución es de 50 a 75 Å, la resolución de la microscopía electrónica por criomicroscopía es de 8 a 20 Å. La técnica

de criomicroscopía no requiere ni fijación ni tñido de la muestra ya que el contraste y la fijación lo da la propiedad del estado congelado. Por lo tanto, en criomicroscopía, las muestras mantienen la arquitectura de las especies en todos sus ejes espaciales permitiendo reconstruir la estructura tridimensional en alta resolución a partir del análisis de miles de imágenes. En la **Figura 10** mostramos el resultado del análisis estadístico de las partículas de menor diámetro observadas por criomicroscopía y su reconstrucción tridimensional. Se promediaron 16000 imágenes conteniendo partículas esféricas de diámetro $\sim 3.5 \pm 0.5$ nm. Del promedio se destacaron cuatro grupos morfológicos predominantes compuestos por una, dos, tres y cuatro esferas (**Figura 10A**). A partir de dicho análisis se reconstruyó la estructura tridimensional de los grupos compuestos por una, dos y tres partículas (ver métodos), las cuales fueron dispuestas en la misma escala del cristal correspondiente a scE2C (**Figura 10B**, fucsia). El diámetro de las partículas observadas coincide con el diámetro de scE2C (42 ± 1 Å), sugiriendo que estaríamos analizando la estructura tridimensional del núcleo, y por lo tanto confirmando la población del mismo de acuerdo a lo observado en los estudios cinéticos. De ser así, la resolución de la reconstrucción tridimensional es suficiente para afirmar que la identidad globular de scE2C se mantiene en el núcleo.

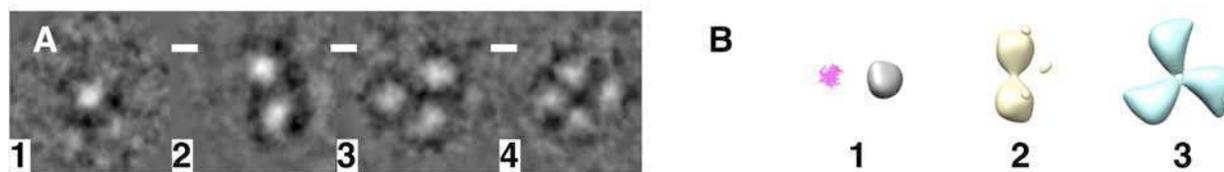


Figura I.10. Representación tridimensional de las imágenes obtenidas por CryoEM. (A) Imágenes promedio resultantes del análisis de 16000 partículas. Se observaron cuatro morfologías mayoritarias, numeradas de acuerdo al número de esferas que la componen. Barra de escala horizontal: 3.5 nm. (B) Reconstrucción tridimensional de las especies observadas compuestas por 1, 2 y 3 esferas. En color fucsia: estructura tridimensional del monómero de scE2C por rayos X (PDB: 2Q79), no mostramos la reconstrucción del grupo 4 dada su baja resolución.

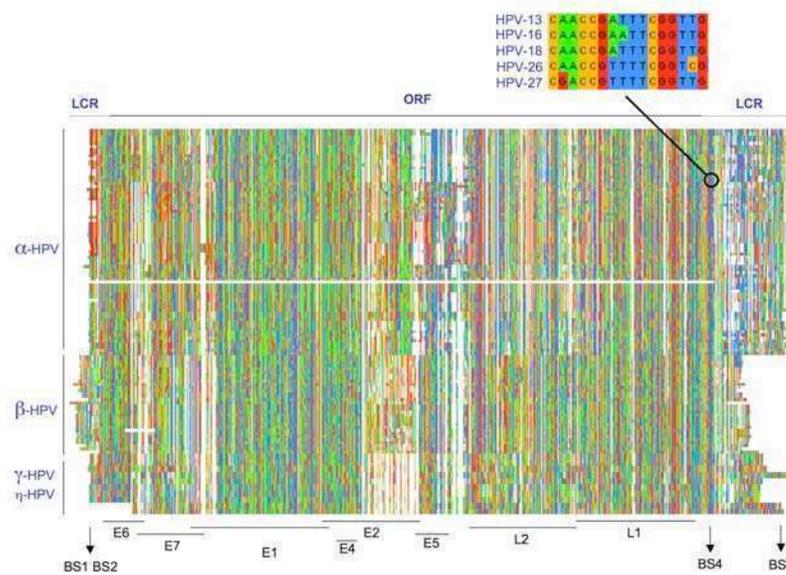
Finalmente, en cuanto al grupo 4 (**Figura 10A**), su reconstrucción tridimensional resultó de muy baja resolución (no mostrado), pero dado que el diámetro promedio corresponde a 10 nm, esta especie entraría dentro de la categoría bola pequeña observada por AFM.

Hemos visto entonces los efectos de la temperatura en la plasticidad conformacional de E2C y abordado su mecanismo de formación. El paisaje morfológico producto de la polimerización de E2C sugiere la presencia de una jerarquía en la aparición de las protofibras amiloideas. En la sección Discusión analizaremos este aspecto resumiendo gráficamente los

resultados presentados. Por otra parte, el ADN específico inhibe notablemente el camino amiloide de E2C. A continuación abordaremos el estudio de la interacción entre E2C y ADN que además de impedir la formación de especies no funcionales de E2C, orquesta el desarrollo del ciclo infeccioso de papilomavirus humano.

Capítulo II

“Análisis estadístico de la discriminación de secuencias específicas de ADN por la proteína E2”



*When it comes to viral evolution,
the median is not the message.*

Stephen Jay Gould por Vincent Racaniello

En la Introducción nos referimos a la jerarquía de afinidades del reconocimiento de E2 a los cuatro sitios presentes en la región reguladora de los genomas de papilomavirus del *genus* alfa y mencionamos que existen diferencias sutiles de jerarquía entre diferentes tipos virales, en particular entre tipos virales de alto y bajo riesgo (Alexander y Phelps 1996; Dell *et al.* 2003). Con el fin de determinar el origen de estas diferencias de jerarquía, a continuación examinamos la conservación de la interacción de todos los tipos de papilomavirus alfa.

Conservación de E2 en la hélice de unión a ADN.

En primer lugar, analizamos la región de E2 involucrada en la interacción con ADN, la hélice $\alpha 1$. La **Figura 1** muestra el complejo del dominio E2C-18 con su ADN diana (Kim *et al.* 2000). Los dominios de E2C de HPV6 y HPV16 forman complejos binarios muy similares (Hegde 2002; Cicero *et al.* 2006; Hooley *et al.* 2006). Las estructuras de los dominios aislados de E2C de HPV-16 u HPV-31 también se encuentran conservadas (Liang *et al.* 1996; Hegde y Androphy 1998; Hegde 2002; Nadra *et al.* 2004). La hélice $\alpha 1$ de reconocimiento de E2 es el mayor determinante para la unión a ADN (**Figura 1**) (Ferreiro *et al.* 2005). El logo de secuencia (Schneider y Stephens 1990) en la **Figura 1** muestra la conservación de los residuos en la hélice en los tipos de alfapapilomavirus en una escala de conservación en el rango de 0 a 4.32 bits. Esta región de la proteína está altamente conservada, como se puede apreciar por la altura de las letras apiladas que se acercan al máximo valor de conservación (Schneider y Stephens 1990). Determinaciones experimentales en nuestro laboratorio indican que la mayoría de los residuos de la hélice contribuyen en más de 0.8 kcal/mol a la energía de unión (Ferreiro *et al.* 2005), y se encuentran muy conservados (notar residuos marcados con asterisco en el logo de la **Figura 1**). En concordancia con la conservación de la estructura en el complejo y la conservación de las secuencias de las proteínas, la especificidad de unión a ADN de E2C es semejante para dominios homólogos distantes, (Alexander y Phelps 1996; Kim *et al.* 2000; Zhang *et al.* 2004) graficado en (Sanchez *et al.* 2008) y en el Capítulo IV de la presente tesis donde distinguimos que cinco dominios E2C discriminan ADN en forma similar. Además, la energía de unión absoluta de los dominios E2C se encuentra conservada en distintos tipos de alfapapilomavirus, (Alexander y Phelps 1996; Hines *et al.* 1998; Kim *et al.* 2000; Dell *et al.* 2003), como estudiaremos en el Capítulo IV.

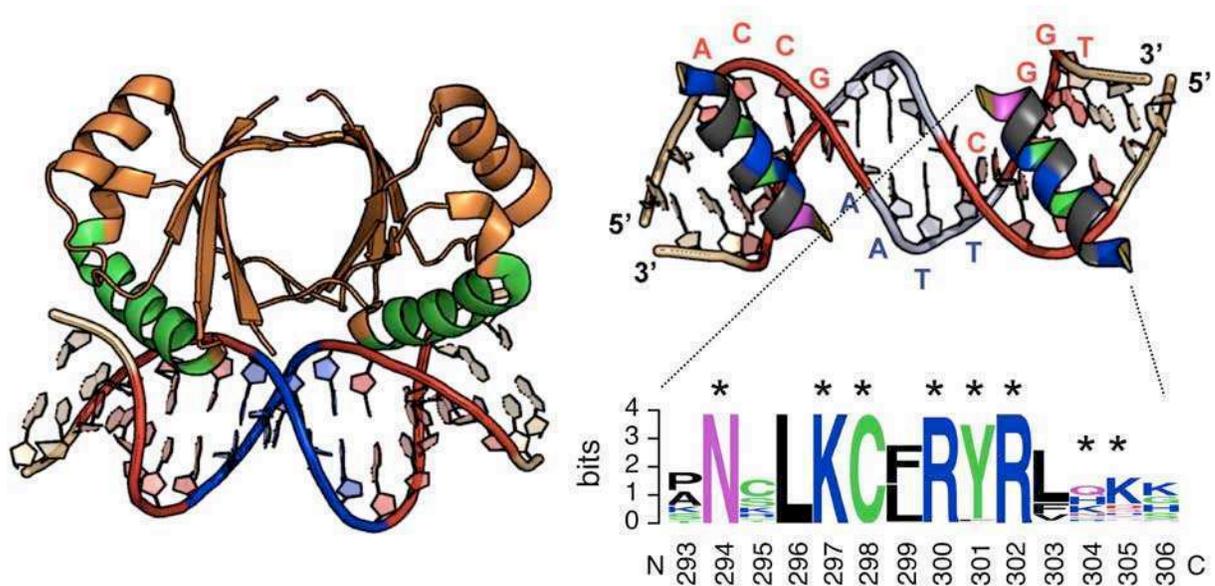


Figura II.1. Características conservadas de la interacción E2-ADN. **Izquierda:** complejo del dominio C-terminal de E2 de HPV tipo 18 con la secuencia idealizada AACCGAATTCGGTT (PDB: 1jj4, (Kim *et al.* 2000)). Las cuatro bases de los dos hemisítios de contacto directo con la proteína están coloreadas en rojo, el espaciador entre cada hemisítio en azul y las bases flanqueantes en blanco. Las hélices $\alpha 1$ de E2C involucradas en el contacto directo con el ADN están coloreadas en verde. **Derecha arriba:** complejo E2C-ADN mostrando para E2C únicamente las hélices $\alpha 1$ las cuales están coloreadas de acuerdo a las propiedades químicas de cada residuo: azul básicos, verde hidrófilos cargados, magenta hidrófilos no cargados y negro hidrófobos. **Derecha abajo:** logo de secuencias (Schneider y Stephens 1990; Crooks *et al.* 2004) correspondiente a la hélice $\alpha 1$ construido en base al alineamiento de todas las E2 de alfapapilomavirus. El asterisco destaca los residuos con contribuciones mayores a 0.8 kcal/mol para la unión (Ferreiro *et al.* 2005).

Entonces, como primera conclusión observamos que las diferencias que se observan en alfapapilomavirus en cuanto a la interacción de E2 con ADN se deben mayormente a diferencias en la secuencia del ADN y no en el dominio de E2C. Dado que la función de cada sitio de ADN de unión de E2 se encuentra conservada para tipos 6, 11, 16, 18 y 31, (Thierry y Howley 1991; Chiang *et al.* 1992; Remm *et al.* 1992; Lu *et al.* 1993; Dong *et al.* 1994; Demeret *et al.* 1995; Sverdrup y Khan 1995; Demeret *et al.* 1997; Rapp *et al.* 1997; Steger y Corbach 1997; Stubenrauch *et al.* 1998; Newhouse y Silverstein 2001; Soeda *et al.* 2006), formulamos la hipótesis de que las diferencias fenotípicas entre los tipos dada por la interacción entre E2 y ADN se debe a diferencias entre los sitios de unión de E2.

Conservación y asimetría de los cuatro sitios de unión.

Hemos considerado para este estudio los sitios de unión de E2 de 73 tipos de alfapapilomavirus: HPV 2a, 2isoC1, 2isoC2, 3, 6, 6a, 6b, 7, 10, 11, 13, 13b, 16, 18, 26, 27, 27b, 28, 29, 30, 31, 32, 33, 34, 35, 35H, 39, 40, 42, 43, 44, 45, 51, 52, 53, 54, 55, 56, 57, 57b, 58, 59,

61, 62cand, 66, 67, 68a, 69, 70, 71, 72, 73, 74subtipo, 77, 81, 82, 82subtipo, 83, 84, 085cand, 086cand, 087cand, 089cand, 090cand, 91, 94, 94korean, 97, 97iso624, 102 y 106, PCPV1 y RHPV1 ((Sanchez *et al.* 2008), material suplementario). Extrajimos todos los sitios de los genomas y los alineamos (ver métodos). Encontramos los cuatro sitios (**Figura 2A**) en todos los genomas, excepto para el BS2 del tipo 94korean. Elegimos mostrar la variabilidad de los sitios de unión de E2 utilizando logos de secuencia (Schneider y Stephens 1990; Crooks *et al.* 2004). Los logos representan el alineamiento de los sitios de ADN en una fila de letras apiladas. La altura de una letra es proporcional al contenido de información en esa posición del alineamiento, el cual puede ser interpretado como una medida de conservación. Para ADN, el rango de valores es entre 0 y dos bits. La altura de las letras dentro de un apilamiento es proporcional a la abundancia de cada base. Las secuencias consenso no muestran la conservación de la secuencia y son limitadas a solo una base por posición (Schneider y Stephens 1990).

Los logos de los cuatro sitios de unión se muestran en la **Figura 2B**. BS1, BS2 y BS4 se encuentran mucho más conservados que BS3. Los logos de todos los sitios cumplen con la secuencia consenso establecida aACCg(A/T)₄cGGTt. Tanto las bases en los dos hemisitios con contactos directos con la proteína (lectura directa (Ferreiro *et al.* 2005)) y las cuatro bases del espaciador (lectura indirecta (Zhang *et al.* 2004)) se encuentran significativamente conservadas. A la fecha de inicio esta tesis doctoral, se creía que todas las bases del espaciador tenían la misma importancia. A pesar de ello, hay un claro gradiente de conservación de la posición 6 más conservada a la posición 9 menos conservada en BS1, BS2 y de la 9 a la 6 en BS4 (**Figura 2B**). Los sitios de unión son generalmente asimétricos, siendo solo 11 (4%) completamente palindrómicos ((Sanchez *et al.* 2008), material suplementario). Dada esta falta de simetría, se deduce que no es equivalente para la unión de E2 si la orientación de un sitio es directa o inversa (Berg y von Hippel 1988; Sarai *et al.* 1989; Raumann *et al.* 1994; Pyles *et al.* 1998; King *et al.* 1999; Winston *et al.* 2000; Kalodimos *et al.* 2002; Ramirez-Carrozzi *et al.* 2003).

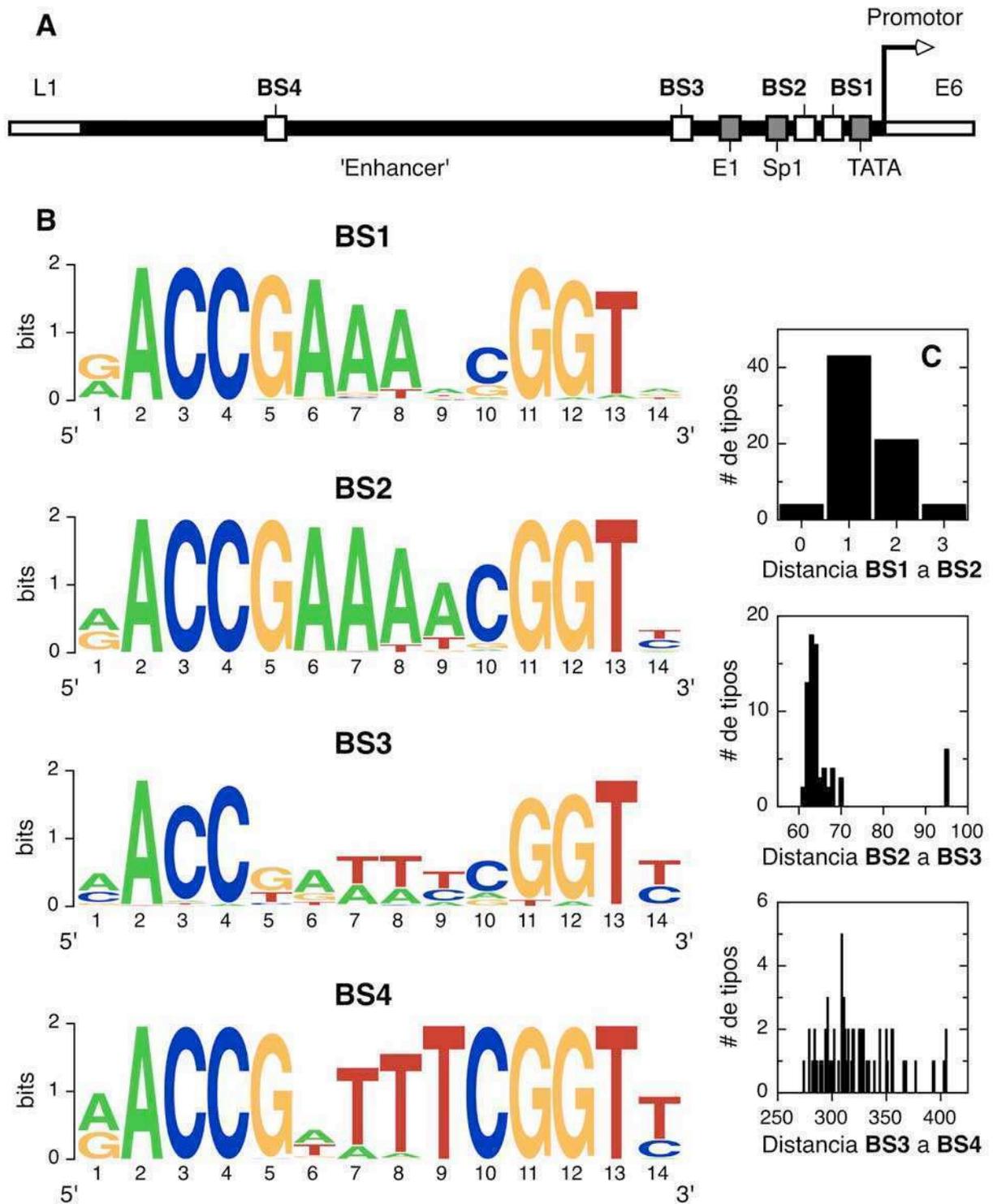


Figura II.2. Sitios de unión en alfapapilomavirus. (A) Vista esquemática de la región reguladora de un genoma prototípico de HPV. Se muestran los marcos de lectura flanqueantes, L1 y E6, el comienzo del promotor temprano y su caja TATA, los cuatro sitios de unión para la proteína E2 (BS1 a 4), el sitio de unión para la proteína viral E1 y para la proteína del huésped Sp1 y la región potenciadora. (B) Logos de secuencia (Schneider y Stephens 1990; Crooks *et al.* 2004) de los cuatro sitios de unión de E2. Los sitios están representados en la dirección 5'-3'. (C) Histograma del número de bases que separan los sitios de unión de E2.

BS1 y BS2 tienen como secuencia espaciadora consenso AAA(A/T) en la dirección 5' a 3', mientras que los sitios BS3 y BS4 poseen dicho consenso únicamente cuando consideramos la secuencia reversa complementaria (**Figura 2B**). Por lo tanto y llamativamente, la orientación de los sitios de unión de E2 parece variar dentro del genoma de HPV. Los resultados en la **Figura 2B** indican que los sitios de E2 difieren en sus niveles de conservación y en su orientación, en línea con la asignación de distintos roles fisiológicos a cada sitio.

A continuación nos propusimos examinar las posiciones relativas de los sitios de unión de E2 en alfafapilomavirus (**Figura 2C**). El número de bases entre BS1 y BS2 se encuentra muy restringido, con distancias que van de 0 a 3 bases y siendo en la mayoría de los casos de 1 o 2 bases. La distancia entre los sitios 2 y 3 es más variable, siendo en ambos casos entre 60 y 70 bases. Curiosamente, la distancia es de 95 bases para 6 tipos, siendo todos ellos cutáneos. Finalmente, entre 275 y 400 bases separan los sitios de unión de E2 BS3 y BS4 sin ningún valor mayoritario. No observamos correlación alguna entre las diferentes distancias (datos no mostrados). Podemos entonces concluir que las distancias de los sitios de E2 evolucionaron en forma independiente y, para el caso de los sitios BS1 a BS3, de forma tal que parece estar restringido por la biología de los alfafapilomavirus.

Posible sitio de metilación conservado dentro del sitio de unión de E2.

La principal presión evolutiva en muchos sitios de ADN de interacción con proteínas, consiste en mantener la energía de la unión por sobre cierto umbral energético (Berg y von Hippel 1987; Berg y von Hippel 1988; Berg y von Hippel 1988). Hemos investigado si este es el caso para los sitios de unión de ADN de la proteína E2 para los tipos de alfafapilomavirus mediante la comparación de los sitios naturales que encontramos en HPV, con un conjunto de sitios seleccionados *in vitro* siendo la afinidad la única restricción (Newhouse y Silverstein 2001). El logo de secuencia en la **Figura 3A**, corresponde a todos los sitios naturales de unión de E2, usando las secuencias reverso complementarias de BS3 y BS4 en el alineamiento (ver arriba). El logo del medio en la **Figura 3A** corresponde al conjunto de sitios resultantes de la amplificación y selección por afinidad *in vitro* de E2C de HPV tipo 51 (Newhouse y Silverstein 2001). La secuencia consenso obtenida *in vitro* para E2C-51 (AACACAAATCGGTT), es reconocida con gran afinidad por E2C de HPV tipo 16 (Newhouse y Silverstein 2001), sugiriendo que estos experimentos pueden ser extrapolados a otros dominios E2C de

alfapapilomavirus. Cualitativamente se observa una coincidencia adecuada entre la mayoría de las posiciones naturales (*in vivo*) y la selección *in vitro*. Sin embargo, la mayor diferencia ocurre para las bases ubicadas en las posiciones 4 y 5.

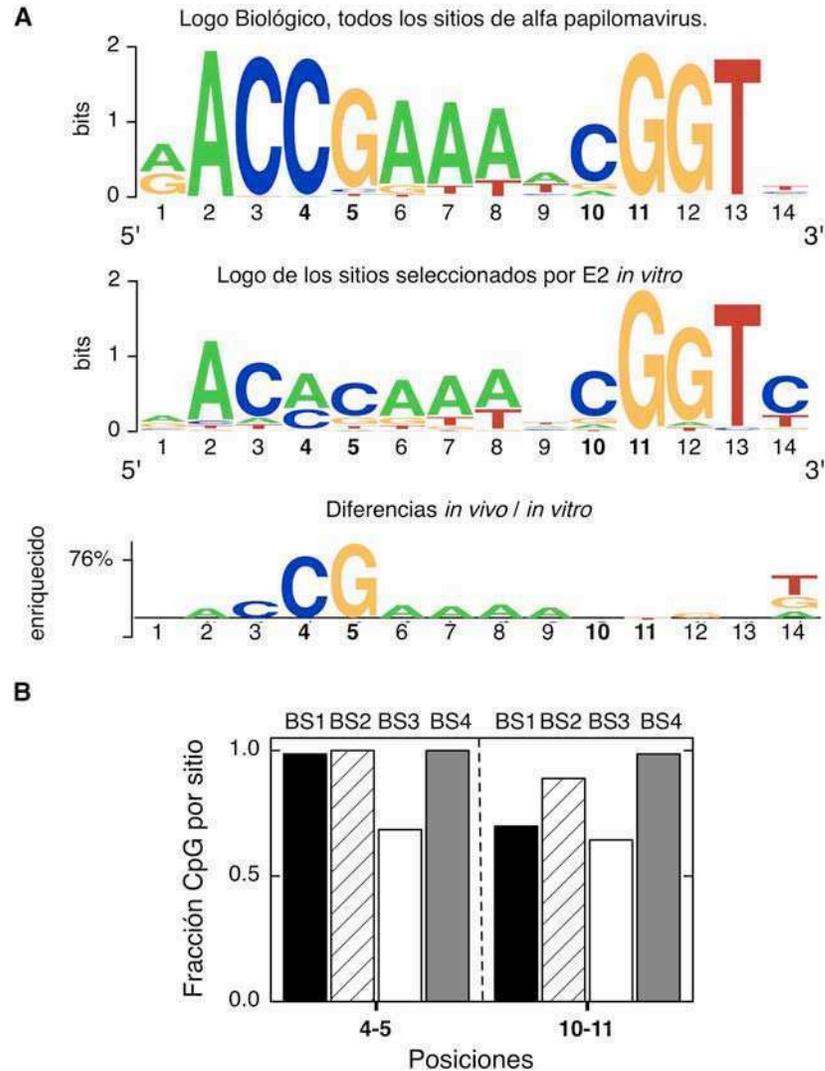


Figura II.3. Influencia de la metilación GC en la evolución de los sitios de unión de E2. (A) Arriba: logo de secuencias para todos los sitios biológicos de unión de E2 de alfapapilomavirus. Al medio: logo de secuencias de experimentos de selección *in vitro* de la unión de E2C de HPV51 a ADN. Abajo: logo diferencial de dos muestras (Vacic *et al.* 2006), siendo los sitios biológicos la muestra y los sitios seleccionados el *in vitro* como contraste. (B) Proporción de la presencia de un sitio putativo de metilación (dinucleótidos CG) en las posiciones 4-5 y 10-11, en cada uno de los cuatro sitios de unión de E2 biológicos.

Podemos visualizar la diferencia entre el *logo in vivo* y el *in vitro* mediante la herramienta *logo de dos muestras* (Vacic *et al.* 2006) (**Figura 3A**, abajo). Este logo toma los sitios seleccionados *in vitro* como referencia y muestra las bases que se encuentran enriquecidas en los sitios naturales. Está claro que la mayor diferencia entre ambos conjuntos de sitios corresponde a

la presencia del dinucleótido CG en las posiciones 4 y 5 en los sitios naturales. Por lo tanto, hemos encontrado una presión evolutiva para la conservación de un sitio de metilación en las posiciones 4 y 5 en el sitio de unión de E2.

Dado que cada sitio de unión tiene efectos diferentes en la replicación de papilomavirus, mantenimiento del genoma y transcripción, buscamos sitios putativos de metilación en las posiciones 4-5 en todos los cuatro sitios de unión de E2. El dinucleótido CG se encuentra presente en la gran mayoría de los BS1, BS2 y BS4 y una mayoría en BS3 (**Figura 3B**). Un segundo dinucleótido CG se encuentra presente en las posiciones 10-11 del sitio de unión de E2, ambos en los logotipos biológicos y los seleccionados (**Figura 3A**). Esto indica que la presencia del segundo sitio CG *in vivo* no es debido a restricciones de afinidad, aunque la evolución para un sitio de metilación no puede ser descartada. Esta metilación putativa en la posición 10-11 se encuentra tan conservada como la 4-5 en BS4 y BS3 y menos conservada en BS1 y BS2 (**Figura 3B**). En conjunto, nuestros resultados sugieren que los cuatro sitios BS diferirían en cuanto al grado de propensión a ser regulados por eventos de metilación.

Predicción computacional de la afinidad de los dominios de unión de E2.

Hemos mostrado que la afinidad de unión es la mayor presión evolutiva en la mayoría de las posiciones de la secuencia de ADN diana de E2 (**Figura 3A**). Se pudo cuantificar también la presión evolutiva para la unión a sitios individuales usando la teoría de mecánica estadística de Berg y von Hippel (Berg y von Hippel 1987; Berg y von Hippel 1988; Berg y von Hippel 1988) (descrito en métodos). Brevemente, la teoría asume que (1) los sitios del ADN diana de una proteína han sido seleccionados por la evolución para mantener la afinidad de unión por sobre un determinado umbral y (2) que las bases del sitio de unión contribuyen en forma independiente y aditiva a la energía libre de unión. Dado estos dos postulados, existe una relación simple entre la frecuencia de ocurrencia de dos bases (i y j) en una posición dada en el repertorio de sitios naturales y la diferencia de su contribución a la energía libre de unión:

$$\Delta\Delta G_{unión}(base_i \rightarrow base_j) = R \cdot T^* \cdot \ln\left(\frac{f(base_i)}{f(base_j)}\right) \quad [1]$$

donde T^* corresponde a una determinada pseudotemperatura relacionada con la tolerancia de la interacción entre la proteína y el ADN a mutaciones y no está relacionada con una temperatura física (Berg y von Hippel 1987; Berg y von Hippel 1988; Berg y von Hippel 1988). La diferencia

en la afinidad entre dos sitios puede ser calculada sumando las diferencias sobre todas las posiciones del sitio:

$$\Delta\Delta G_{\text{unión}}(\text{sitio}_i \rightarrow \text{sitio}_j) = \sum_{\text{posiciones}} \Delta\Delta G_{\text{unión}}(\text{base}_i \rightarrow \text{base}_j) \quad [2]$$

Esta teoría ha sido exitosamente aplicada a predicciones cuantitativas de constantes relativas de unión en varias interacciones proteína-ADN (Berg y von Hippel 1987; Berg y von Hippel 1988; Berg y von Hippel 1988).

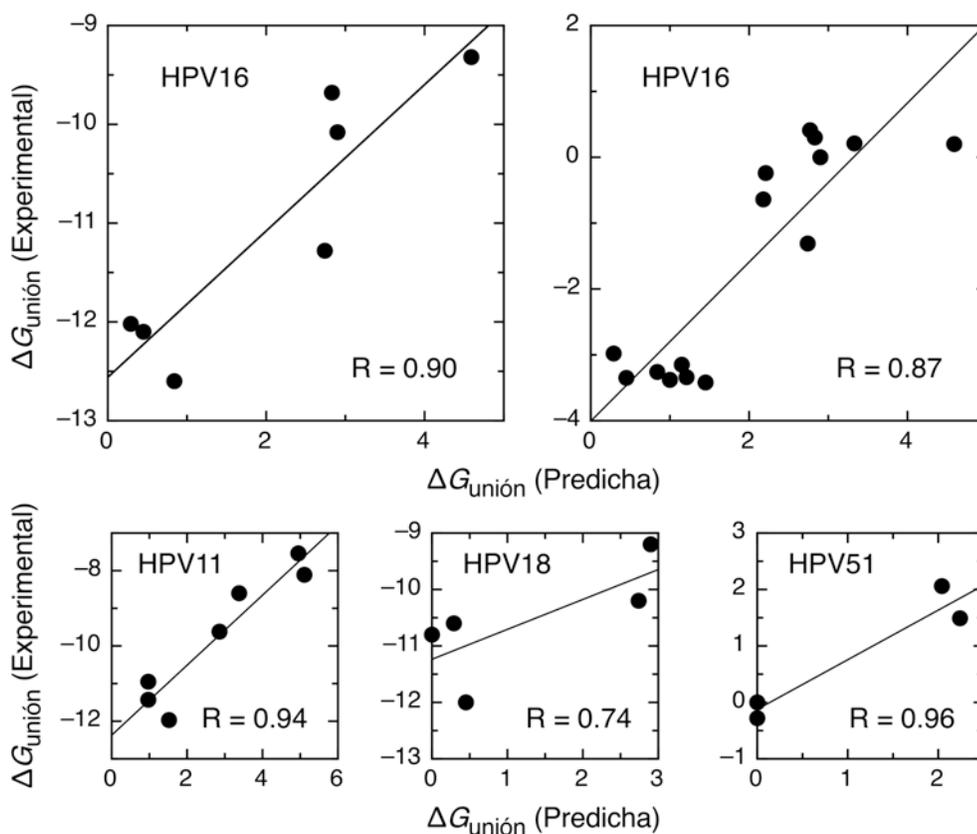


Figura II.4. Correlación entre energías libres predichas y observadas de los complejos E2-ADN con las proteínas E2 de HPV tipo 11, 16, 18 y 51 (Alexander y Phelps 1996; Hines *et al.* 1998; Kim *et al.* 2000; Newhouse y Silverstein 2001; Zhang *et al.* 2004). Las unidades en todos los casos corresponden a kcal/mol. La energía de unión a la secuencia consenso fue arbitrariamente llevada a cero. El número total de puntos es 38, la desviación estándar entre observado y predicho es 0.64 kcal/mol o 2.9 veces en K_D .

Decidimos evaluar si la teoría de Berg-von Hippel se mantiene para la interacción entre E2C y sus sitios de unión diana. Para eso extrajimos de la literatura cinco grupos de energías libres para la unión entre E2C y los dominios E2C de tipo 11, 16, 18 y 51 y de cuatro a quince sitios distintos para cada dominio (Alexander y Phelps 1996; Hines *et al.* 1998; Kim *et al.* 2000; Newhouse y Silverstein 2001; Zhang *et al.* 2004) (**Figura 4**). Posteriormente calculamos la

energía de unión relativa a la secuencia consenso para todos los sitios de alfapapilomavirus, utilizando el alineamiento de todos los sitios naturales de E2, representada en la **Figura 3A** (ver métodos).

La correlación entre datos experimentales y las energías de unión calculadas para los cinco grupos obtenidos de la literatura incluyendo cuatro dominios diferentes de E2C se muestran en la **Figura 4**. Los valores R de correlación se encuentran en el rango de 0.74 a 0.96. El desvío estándar entre los sitios observados y los predichos es 0.64 kcal/mol para 38 puntos, o 2.9 veces expresado en K_D . El rendimiento de nuestro método basado únicamente en secuencias se encuentra cerca del desvío estándar de 0.57 kcal/mol (16 puntos) obtenido para el dominio E2C de HPV16 utilizando un modelo basado en estructura (Zhang *et al.* 2004).

Los experimentos encontrados en literatura validan la utilización de la teoría de Berg-von Hippel para calcular las energías correspondientes la unión de todos los sitios de alfapapilomavirus, aún tratándose de experimentos realizados en distintas condiciones y laboratorios.

Seis jerarquías de afinidad prototípicas en alfapapilomavirus.

Luego de analizar la interacción entre E2C-16 y sus cuatro sitios blanco en solución mediante ITC, retomamos el análisis estadístico de secuencia. Hemos usado la teoría de Berg-von Hippel para calcular la afinidad de E2C para todos los sitios en alfapapilomavirus (explicada en métodos). BS4 es el sitio con mayor afinidad, solo 0.61 ± 0.53 kcal/mol (promedio \pm desvío estándar) peor que el consenso. Es seguido del sitio BS2 (0.91 ± 0.69 kcal/mol), BS1 (1.46 ± 1.47 kcal/mol) y BS3 (1.94 ± 1.60 kcal/mol). Esta jerarquía coincide cualitativamente con los niveles de conservación del logo para cada sitio (**Figura 2B**).

Una inspección manual de todas las jerarquías de afinidad, reveló la existencia de varios grupos de tipos muy similares. Hemos agrupado todos los tipos virales en grupos jerárquicos utilizando el algoritmo de medias-k (descrito en métodos). A través de este procedimiento, cada tipo viral es representado por sus cuatro energías de unión predichas. Los tipos son agrupados de forma tal de minimizar la suma de todas las diferencias entre las energías predichas y los grupos promedio. Por ejemplo, agrupar los datos en menos de seis grupos produjo grupos con jerarquías muy diferentes, por lo tanto cinco o menos grupos no fueron suficientes para describir los datos de manera adecuada (datos no mostrados). Pero al agrupar los datos en más de seis grupos

produjo grupos redundantes, indicando que estos debían estar en un mismo conjunto (datos no mostrados).

El resultado de agrupar en seis grupos se muestra en la **Figura 5**. 33 de los 72 tipos pertenecen al primer grupo (**Figura 5A**), en la cual BS1, BS2 y BS4 tienen alta afinidad y la unión a BS3 es claramente más débil. Para el segundo grupo mayoritario, (compuesto por 20 tipos, **Figura 5B**), todos los BS poseen una energía de unión muy cercana a la de la secuencia consenso. En el tercer grupo mayoritario (9 tipos, **Figura 5C**), BS2 y BS4 poseen una fuerte energía de unión predicha y BS1 y BS3 poseen menor afinidad por E2C. Cuatro tipos pertenecen a un cuarto grupo (**Figura 5D**), en la cual BS2 y se predicen como de alta afinidad, BS1 de débil unión y BS3 de unión marginal. En los tres tipos del quinto grupo, BS1, BS3 y BS4 tienen alta afinidad y BS2 una afinidad débil (**Figura 5E**). Finalmente, en los tres tipos del sexto grupo (**Figura 5F**), BS2 y BS4 presentan una buena afinidad predicha, BS3 una afinidad débil y BS1 posee una afinidad marginal por la proteína.

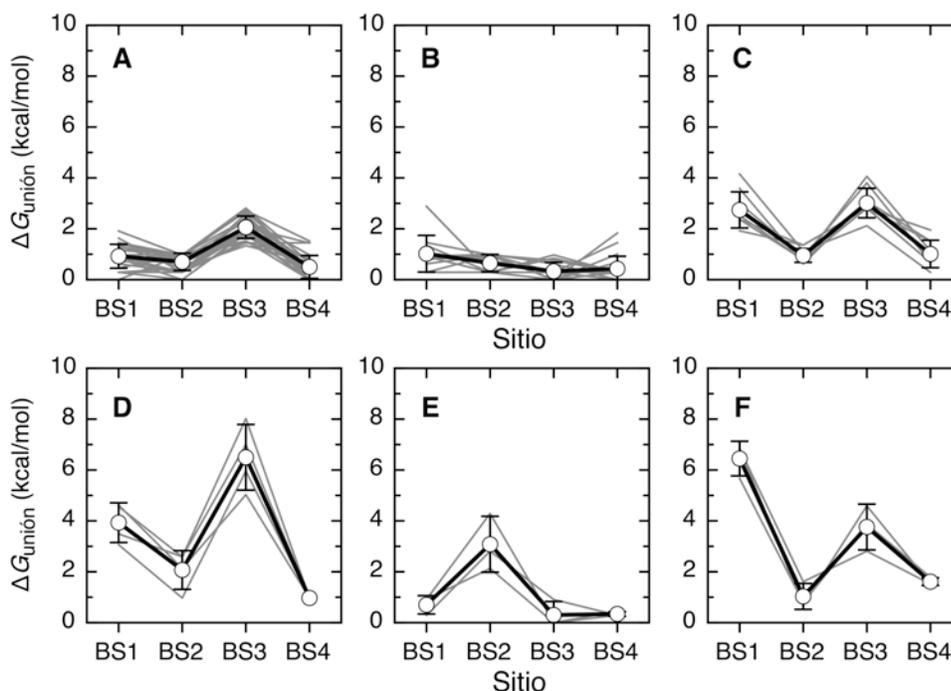


Figura II.5. Seis clases de jerarquías de afinidad de unión para la interacción E2-ADN. Para cada grupo de tipos representamos la afinidad predicha promedio (línea gruesa negra) y la desviación estándar para cada sitio y los valores para cada tipo (líneas finas grises). Los tipos fueron agrupados utilizando el algoritmo de medias-k (métodos).

Nuestros resultados concuerdan con el conocimiento actual del rol de cada sitio. En todas las jerarquías, la afinidad de BS4 es por lo menos tan alta como para los otros sitios (**Figura 5**),

asegurando que los genes tempranos del virus se transcriban (Dong *et al.* 1994; Rapp *et al.* 1997; Steger y Corbach 1997). BS1, BS2 o ambos poseen también una buena afinidad en todas las jerarquías (**Figura 5**), garantizando la replicación viral y la represión de la transcripción de oncogenes (Thierry y Howley 1991; Chiang *et al.* 1992; Remm *et al.* 1992; Lu *et al.* 1993; Dong *et al.* 1994; Demeret *et al.* 1995; Sverdrup y Khan 1995; Demeret *et al.* 1997; Rapp *et al.* 1997; Steger y Corbach 1997; Stubenrauch *et al.* 1998; Newhouse y Silverstein 2001; Soeda *et al.* 2006). BS3, que posee un rol secundario tanto en replicación como en transcripción, posee una afinidad predicha menor que los otros tres sitios en las jerarquías más abundantemente encontradas (**Figura 5**).

Asociación entre las propiedades moleculares y epidemiológicas.

Los tipos de alfafapilomavirus son comúnmente clasificados como de alto-riesgo (en total diecinueve tipos: 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68a, 73, 82, 82subtipo) o bajo-riesgo (en total trece tipos: 6, 6a, 6b, 11, 40, 42, 43, 44, 54, 61, 70, 72 y 81) de acuerdo con la relación de las probabilidades de desarrollar cáncer cervical en pacientes infectados con un tipo y en aquellos pacientes HPV-negativos (Munoz *et al.* 2003). El *genus* alfa incluye todos los tipos genitales pero además incluye tipos que infectan tejidos cutáneos (en total doce tipos: 2, 2a, 2isoc2, 3, 7, 10, 27, 27b, 28, 29, 94 y 94korean). Analizamos si el comportamiento epidemiológico de alto-riesgo, bajo-riesgo y tipos cutáneos se encuentra fenomenológicamente asociado con las propiedades moleculares de la interacción E2C-ADN expuestas en el presente estudio de secuencias. Para ello, utilizamos el modelo probabilístico de distribución hipergeométrica (detallada en métodos). Por ejemplo, en el caso de la distancia entre los sitios BS1 y BS2 observamos que es de dos bases en 43 de los 72 tipos del *genus* alfa, de los cuales 10 tipos son de alto-riesgo. Entonces, el modelo de distribución hipergeométrica calcula la probabilidad de tomar 10 o más tipos de alto-riesgo al elegir al azar 43 tipos de una 'bolsa' con 72 tipos distintos correspondientes a toda la población de alfafapilomavirus (Rivals *et al.* 2007). Si el resultado de esta probabilidad es menor que la elegida con un grado de significancia del 0.05, concluimos que el tipo de alto-riesgo se encuentra asociado a una distancia de dos bases entre BS1 y BS2. Por lo tanto, que 10 tipos de alto riesgo presenten la propiedad de dicha distancia, no es un evento azaroso si no que es una propiedad molecular característica de la epidemiología de alto riesgo.

Tabla II.1. Asociación entre propiedades moleculares y epidemiológicas.

Propiedades moleculares		Propiedades epidemiológicas						
		Alto riesgo (19)		Bajo riesgo (13)		Cutáneos (12)		
		n	Valor p	n	Valor p	n	Valor p	
Afinidad jerárquica	 (33)	7	>0.05	5	>0.05	5	>0.05	
	 (20)	11	1.2·10⁻³ (+)	4	>0.05	1	>0.05	
	 (09)	0	>0.05	0	>0.05	5	4.9·10⁻³ (+)	
	 (04)	0	>0.05	3	1.7·10⁻² (+)	0	>0.05	
	 (03)	1	>0.05	1	>0.05	0	>0.05	
	 (03)	0	>0.05	0	>0.05	1	>0.05	
Defecto de metilación	BS1 {	Posición 4 (1)	1	>0.05	0	>0.05	0	>0.05
		Posición 10 (22)	0	3.6·10⁻³ (-)	3	>0.05	7	2.9·10⁻² (+)
	BS2 {	Posición 4 (0)	0	>0.05	0	>0.05	0	>0.05
		Posición 10 (8)	7	2.3·10⁻⁴ (+)	0	>0.05	0	>0.05
	BS3 {	Posición 4 (23)	14	1.4·10⁻⁵ (+)	2	>0.05	0	6.0·10⁻³ (-)
		Posición 10 (26)	1	6.4·10⁻⁴ (-)	2	>0.05	12	6.3·10⁻⁷ (+)
	BS4 {	Posición 4 (0)	0	>0.05	0	>0.05	0	>0.05
		Posición 10 (1)	0	>0.05	0	>0.05	1	>0.05
Distancia intersitio	d ₁₂ = 0 (4)	1	>0.05	1	>0.05	0	>0.05	
	d ₁₂ = 1 (43)	5	7.2·10⁻⁴ (-)	10	>0.05	12	1.0·10⁻³ (+)	
	d ₁₂ = 2 (21)	10	1.1·10⁻² (+)	2	>0.05	0	1.0·10⁻² (-)	
	d ₁₂ = 3 (4)	3	>0.05	0	>0.05	0	>0.05	
	d ₂₃ = 95 (6)	0	>0.05	0	>0.05	6	5.9·10⁻⁶ (+)	

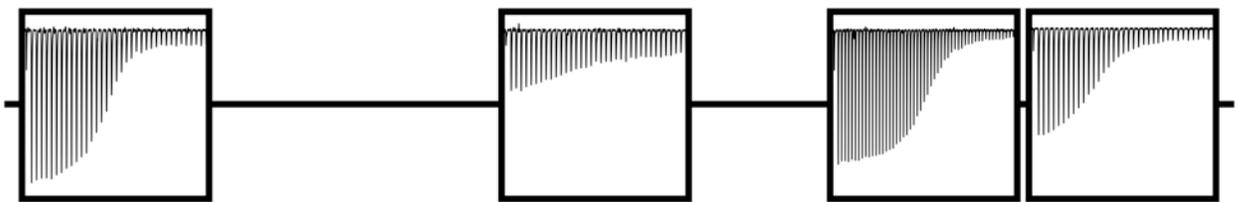
Las propiedades epidemiológicas se muestran en las columnas y las propiedades moleculares en las filas, con el número de tipos entre paréntesis. Para una dada combinación de propiedades, indicamos entre paréntesis, el número de tipos (n) incluidos y la probabilidad (p) que la observación ocurra al azar. Valores significativos de p se destacan en tipografía negra. Los símbolos de adición y de resta indican las combinaciones de propiedades epidemiológicas y moleculares que ocurren con mayor o menor frecuencia que al azar.

Los resultados para la asociación entre fenotipos y propiedades moleculares se muestran en la **Tabla 1**. Los tipos de alto-riesgo se encuentran asociados con la jerarquía de afinidades de los sitios de unión de E2 de la **Figura 5B**, los tipos de bajo-riesgo con la afinidad de jerarquías de la **Figura 5D**. Dado que la mayoría de los sitios de unión de E2 poseen dos sitios de metilación putativos en las posiciones 4-5 y 10-11 (**Figura 3**), observamos la asociación del comportamiento epidemiológico en ausencia de sitios de metilación (**Tabla 1**, defecto de metilación). Los tipos de alto-riesgo se encuentran asociados con la falta de bases ‘metilables’ en la posición 10-11 del BS2 y en las posiciones 4-5 de BS3. Por otro lado los tipos de alto-riesgo se encuentran asociados con la presencia de bases metilables en la posición 10-11 en los sitios BS1 y BS3.

Finalmente, apreciamos que los tipos cutáneos se encuentran positivamente asociados con la falta de bases CpG en las posiciones 10-11 de los sitios BS1 y BS3 y negativamente asociados con la falta de bases metilables en la posición 4-5 del sitio BS3. Respecto al número de bases entre sitios, los tipos de alto-riesgo se encuentran positivamente asociados con una distancia de dos bases entre los sitios BS1 y BS2 y negativamente asociados con la presencia de una única base entre los sitios BS1 y BS2, mientras que exactamente lo opuesto ocurre para los tipos cutáneos (**Tabla 1**). Los sitios BS1 y BS2 dispuestos en tándem (DBS, por doble sitio de unión) forman parte entonces de un bloque funcional clave en la epidemiología del virus como estudiaremos en profundidad en el cuarto Capítulo. Finalmente, los tipos cutáneos se encuentran positivamente asociados a una distancia de 95 bases entre los sitios BS2 y BS3 (**Figura 2 y Tabla 1**). No observamos una asociación estadística con probabilidad menor que 0.05 entre propiedades epidemiológicas y la distancia entre los sitios BS3 y BS4 (datos no mostrados). En conjunto, logramos asociar la propiedad epidemiológica alto riesgo con seis propiedades moleculares diferentes, la propiedad epidemiológica bajo con una sola característica molecular y los tipos de HPV que infectan tejido cutáneo con siete propiedades moleculares (**Tabla 1**). Los estudios epidemiológicos progresan velozmente con el avance tecnológico de la biología molecular, es por ello que tanto el número de genomas de distintos tipos que conforman cada *genus* como el número de *genus* existentes en la familia de papilomavirus aumenta año a año (Bernard *et al.* 2010). Por lo tanto, el grado y número de asociaciones, entre las características moleculares y las epidemiologías del virus descritas, pueden aumentar con el correr del tiempo.

Capítulo III

“Origen termodinámico de la discriminación de secuencias específicas de ADN por la proteína E2”



A line is a dot that went for a walk.

Paul Klee

En este Capítulo usamos la calorimetría de titulación isotérmica (ITC) para estudiar las bases termodinámicas de la unión de E2C-16 al ADN. La disección de la energía libre de unión en sus componentes entálpico y entrópico nos va a permitir examinar varias cuestiones planteadas en trabajos previos sobre este complejo, incluyendo el Capítulo anterior.

Dependencia de la concentración de fosfato en la unión de E2C-16 y BS2.

Antes de analizar la unión de E2C-16 al ADN por ITC, exploramos las condiciones de medición utilizando el sitio de unión BS2. La unión de E2C-16 al sitio BS2 ha sido ampliamente estudiada mediante ensayos espectroscópicos en diferentes entornos químicos (Ferreiro *et al.* 2000; Ferreiro 2003; Dellarole 2005; Ferreiro *et al.* 2005; Cicero *et al.* 2006; Dellarole *et al.* 2007; Ferreiro *et al.* 2008; Eliseo *et al.* 2009; Sanchez *et al.* 2010). Los extensos antecedentes de ensayos de unión de E2C-16 a BS2 facilitaron la elección de las condiciones adecuadas. Dichas condiciones deben garantizar que la estequiometría de la unión sea 1:1 (Ferreiro *et al.* 2000) y que la cuantificación de la energía de unión sea válida en todas las titulaciones que buscamos medir (Wiseman *et al.* 1989). Como tampón de medición elegimos el fosfato de sodio el cual por su carga y composición es un compuesto semejante al ADN, mimetizando el entorno fisiológico de E2. El fosfato de sodio le confiere una mayor estabilidad al dominio E2C-16 (Mok *et al.* 1996) y por lo tanto disminuye la fracción incompetente a lo largo de las titulaciones lentas de ITC. El fosfato de sodio fue utilizado para estudiar mediante ensayos espectroscópicos la dependencia de la constante de disociación con la fuerza iónica (Dellarole 2005; Dellarole *et al.* 2007) y en ITC para comparar la entalpía de unión en dominios monomerizados de E2C (Dellarole *et al.* 2007).

A continuación medimos el reconocimiento de E2C-16 a BS2 a diferentes concentraciones de fosfato de sodio con el fin de establecer las condiciones óptimas de medición por ITC de la interacción E2C-ADN a lo largo de la presente tesis. En la **Figura 1A** representamos la unión de E2C-16 a BS2 medida a concentraciones de fosfato de 50 y 75 mM. Dichas concentraciones corresponden a una fuerza iónica inferior a 0.15 M, observamos que ambas curvas presentan dos transiciones antes de llegar al *plateau* final. Este comportamiento reporta más de un único evento de unión en esas condiciones. Medimos la estequiometría a 75 mM de fosfato (**Figura 1B**, puntos vacíos) y observamos que la reacción presenta una molaridad

cercana a 2:1, mientras que la estequiometría medida a 200 mM de fosfato (**Figura 1B**, puntos negros) es del orden de 1:1. Por lo tanto, las mediciones a concentraciones de fosfato inferiores a 75 mM (**Figura 1A**) presentan un evento de unión adicional, en excelente acuerdo con observaciones previas llevadas a cabo en nuestro laboratorio medidas por fluorescencia en otro tampón químico de medición (Ferreiro *et al.* 2000). El evento adicional de unión ocurre a baja fuerza iónica indicando que es de una unión principalmente electrostática e inespecífica y la cual se observó anteriormente que produce la formación de agregados amorfos (Ferreiro *et al.* 2000). Concluimos entonces que el evento adicional presenta un calor exotérmico observable por ITC y que la interacción E2C-DNA debe ser medida a concentraciones de fosfato superiores a 75 mM.

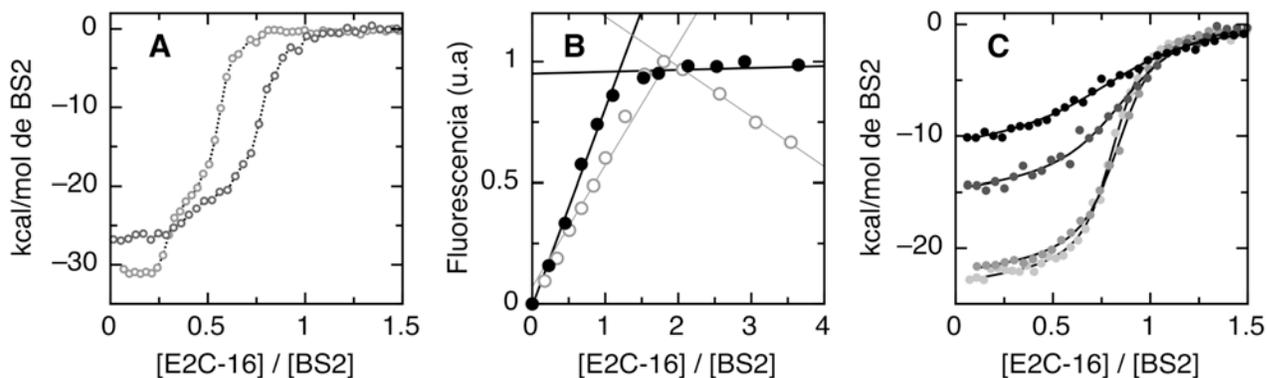


Figura III.1. Unión de E2C-16 a BS2 seguida por ITC en función de la concentración de fosfato de sodio. (A) Datos de ITC integrados y normalizados por concentración correspondientes al complejo entre E2C-BS2 medidos a 50 mM (puntos con línea gris claro) y 75 mM (puntos con línea gris oscuro) de fosfato de sodio. (B) Datos de titulaciones isotérmicas seguidas por fluorescencia en condiciones estequiométricas medidos a 75 mM (gris) y 200 mM (negro) de fosfato de sodio, la intersección de las rectas indica la relación molar en cada titulación. (C) Datos de ITC integrados y normalizados por concentración, donde la línea sólida corresponde al ajuste del modelo de unión 1:1 a los datos medidos a 150 mM, 200 mM 300 mM y 400 mM de fosfato de sodio, (representados en escala de grises, gris claro baja concentración de fosfato de sodio y negro alta concentración de fosfato de sodio). Todos los experimentos fueron medidos a 25°C en diferentes concentraciones de fosfato de sodio pH 7 y 0.2 mM DTT.

A diferencia de lo observado a baja fuerza iónica, los experimentos de titulación realizados en el rango de 100 mM a 400 mM de fosfato de sodio producen curvas de ITC con una única transición (**Figura 1C**). El modelo de un sitio de unión esperado para la interacción E2C-BS2 (Ferreiro *et al.* 2000; Ferreiro 2003; Dellarole 2005; Ferreiro *et al.* 2005; Cicero *et al.* 2006; Dellarole *et al.* 2007; Ferreiro *et al.* 2008; Eliseo *et al.* 2009; Sanchez *et al.* 2010), pudo ser ajustado a los datos de ITC normalizados por concentración (**Figura 1C**, líneas negras). En la **Tabla 1** presentamos los parámetros obtenidos del ajuste a distintas condiciones de fuerza iónica. En todos los casos la estequiometría de unión es cercana a 1 y el valor *C* (detallado en métodos)

es superior a 10 e inferior a 500, garantizando la bondad del ajuste del modelo 1:1 (proteína:ADN) a los datos (Wiseman *et al.* 1989). Es de esperar que en una reacción 1:1 el orden de la titulación no influya en el resultado del ajuste. Medimos calorimetrías de titulación isotérmica en 200 mM de fosfato pH 7, en ambas direcciones: ya sea titulando BS2 sobre E2C, titulaciones ‘directas’ (**Figura 2, izquierda**), como inyectar E2C sobre BS2, titulaciones ‘reversas’ (**Figura 2, derecha**). En efecto y reforzando el modelo de unión 1:1, las estequiometrías y los parámetros termodinámicos de las titulaciones directas y reversas en este sistema binario son similares, tanto para los ajustes individuales y como para el ajuste global de los datos (**Tabla 1**). En sistemas de mayor complejidad, variar el orden de la titulación sí influye en el resultado del ajuste como describimos en el último Capítulo, **Figuras V.3 y V.4**.

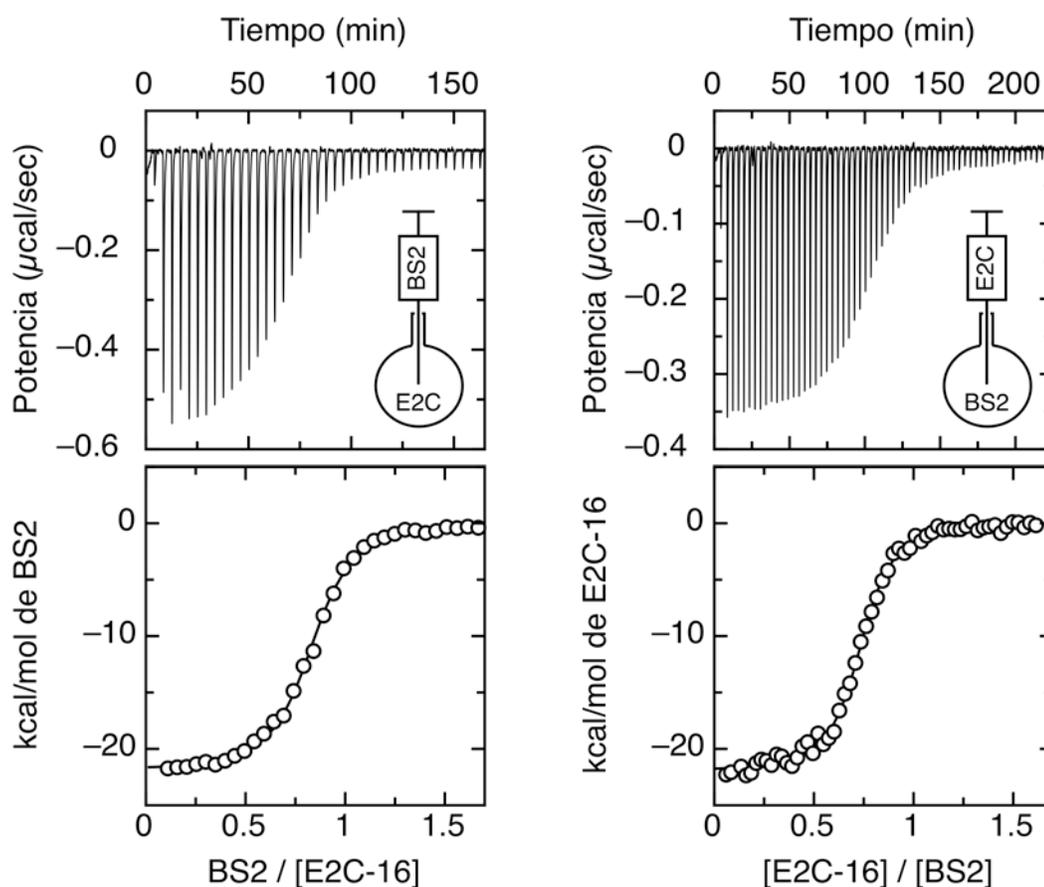


Figura III.2. Titulaciones directas y reversas para la formación de complejo binario entre E2C y BS2. Datos de ITC crudos (paneles de arriba) e integrados y normalizados por concentración (paneles de abajo). **Izquierda:** titulación directa: inyectando el BS2 a una celda conteniendo la proteína E2C. **Derecha:** titulación reversa, inyectando E2C a una celda conteniendo el ADN. Las líneas en los paneles de abajo se corresponden con el ajuste global (Houtman *et al.* 2007) del modelo de unión 1:1 a los datos entre ambas titulaciones directas y reversas. Los experimentos fueron realizados en el tampón químico 200 mM de fosfato de sodio pH 7, 0.2 mM de DTT a 25°C.

Tabla III.1. Parámetros de la unión de E2C-16 a BS2 en función de la fuerza iónica.

[Fosfato] (mM)	$FI^{\#}$ (M)	ΔH (kcal/mol)	K_D (nM)	ΔG (kcal/mol)	$-T \cdot \Delta S$ (kcal/mol)	N	C
100	0.186	-23.6±0.1	76±4	-9.71±0.04	13.9±0.1	0.7±0.1	115±5
150	0.281	-22.8±0.1	107±7	-9.51±0.04	13.3±0.1	0.8±0.1	74±2
200 _{directa}	0.378	-22.3±0.1	130±10	-9.39±0.09	12.9±0.1	0.8±0.1	65±1
200 _{reversa}	0.378	-22.9±0.1	73±5	-9.73±0.10	13.2±0.1	0.7±0.1	81±2
200 _{global}	0.378	-22.7±0.4	90±40	-9.61±0.24	13.1±0.2	0.8±0.1	83±24
300	0.572	-14.9±0.2	330±18	-8.84±0.07	6.1±0.2	0.8±0.1	26±1
400	0.767	-11.1±0.2	764±74	-8.34±0.06	2.8±0.2	0.8±0.1	12±1

[#] Fuerza iónica calculada en base a la aplicación en línea: (Beynon 1996). El ajuste a un modelo de unión 1:1 de los datos genera el cambio en entalpía (ΔH), la constante de disociación (K_D) y la estequiometría de la reacción (N). $\Delta G = -RT \cdot \ln(K_D) = \Delta H - T \cdot \Delta S$ y $C = N \cdot K_D \cdot [\text{ligando}]$ (Wiseman *et al.* 1989). Los errores fueron propagados a partir de ΔH , K_D y N, salvo para ajuste global a 200 mM de fosfato que corresponde al desvío estándar de parámetros de titulaciones directas y reversas.

Al analizar la disección termodinámica de la dependencia de la interacción con la concentración de fosfato, observamos que el aumento de la fuerza iónica de 0.186 M a 0.767 M genera cambios notables en la magnitud de ΔH y $-T\Delta S$ (de ~12 kcal/mol, **Tabla 1**) aunque en todos los casos la interacción E2C-BS2 se mantiene entálpicamente favorable (valores de ΔH negativos, **Figura 3A**, barras grises) y entrópicamente desfavorable (valores de $-T\Delta S$ positivos **Figura 3A**, barras negras). Mientras que el complejo se debilita en 2 kcal/mol (ΔG), correspondiente a un orden de magnitud en constante de disociación (K_D) (**Tabla 1**), observamos que el cambio mayor en ΔG , ΔH y $-T\Delta S$ se presenta al superar la concentración en fuerza iónica (FI) de 0.5 M (**Tabla 1** y **Figura 3A**). Al graficar el ΔH medido en función del logaritmo de la concentración de fosfato (**Figura 3B**), dicho quiebre se evidencia en dos dependencias lineales cuya pendiente es significativamente mayor para los puntos medidos a $FI > 0.5M$ que para los puntos medidos a $FI < 0.5M$. A $FI < 0.5 M$, la actividad del agua se la considera nula (ecuación [10 y 11] detalladas en métodos (Record *et al.* 1978; Record *et al.* 1991)), por lo tanto consideramos que el cambio observado corresponde a un cambio en la actividad del agua en la diferencia energética entre las moléculas E2C-16 y BS2 libres y unidas.

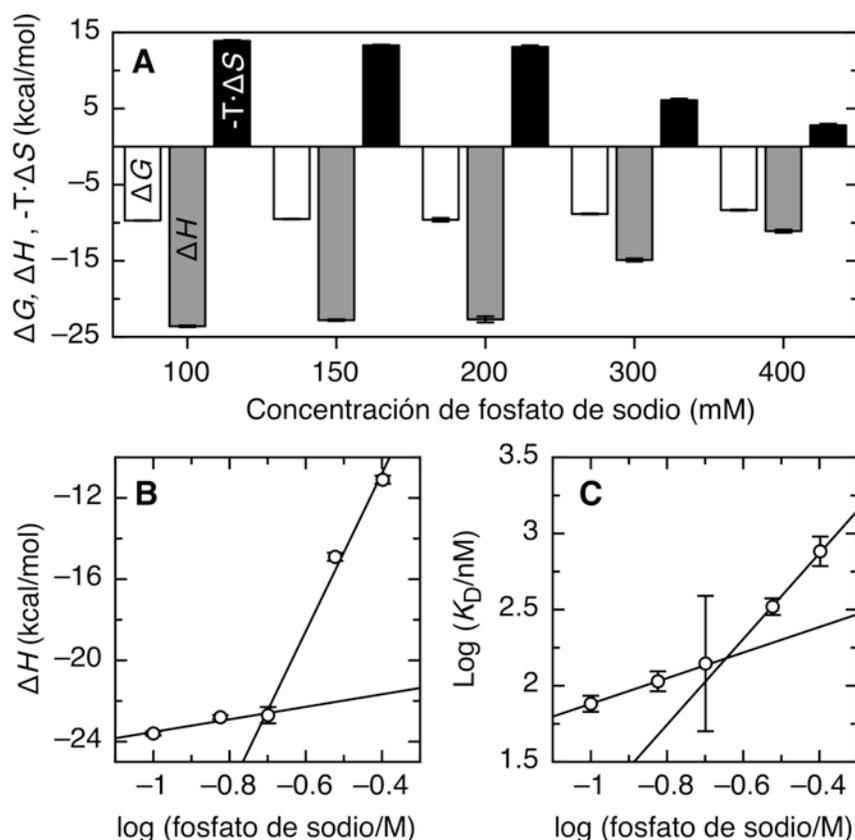


Figura III.3. Análisis de la dependencia de la interacción E2C-BS2 con la concentración de fosfato de sodio. (A) Disección termodinámica de la unión de E2C-16 a ADN seguida por ITC en diferentes concentraciones de fosfato. (B) Entalpía de la reacción observada para la interacción E2C-BS2 a diferentes en función del logaritmo de la concentraciones de fosfato. Las líneas representan un análisis de regresión lineal de los datos medidos a baja concentración de fosfato, pendiente 4.2 ± 0.7 , y los datos medidos a alta fuerza iónica, pendiente 36.8 ± 1.4 . (C) Gráfico log-log de la dependencia de K_D con la concentración de fosfato. Las líneas representan un análisis de regresión lineal de los datos medidos a baja concentración de fosfato, cuya pendiente es $SK_{\text{obs}} = 0.84 \pm 0.46$, y los datos medidos a alta fuerza iónica, cuya pendiente es $SK_{\text{obs}} = 2.8 \pm 0.8$, ecuaciones [10] y [11] en métodos.

La reducción significativa en la afinidad al aumentar la fuerza iónica es consistente con uniones secuencia específica (Record *et al.* 1978; Record *et al.* 1991), y representa la liberación de cationes de la molécula de ADN. Dicha dependencia de la afinidad con la fuerza iónica responde en forma lineal en un gráfico log-log (**Figura 3C**) cuya pendiente representa el coeficiente SK_{obs} , correspondiente a la cantidad de cationes desplazados del ADN en estado libre (ecuación [10] y [11] métodos, (Record *et al.* 1978; Record *et al.* 1991)). Nuevamente observamos dos pendientes en la relación log-log, a concentraciones de $FI < 0.5$ M cuya pendiente es $SK_{\text{obs}} = 0.84 \pm 0.46$ y a concentraciones de $FI > 0.5$ M cuya pendiente es de $SK_{\text{obs}} = 2.8 \pm 0.46$ M. Este cambio en la pendiente corresponde a un cambio en la actividad del agua en la reacción (Record *et al.* 1978; Record *et al.* 1991) que se evidencia a partir de concentraciones de fosfato superiores

a 200 mM. Por lo tanto, concluimos por un lado que la interacción en el amortiguador fosfato de sodio, genera únicamente la liberación de 1 contraíón de la interfaz del ADN, de acuerdo con una interfaz del complejo ‘húmeda’ y altamente dinámica (Lima *et al.* 2004; Ferreiro *et al.* 2005), a diferencia de interfaces ‘secas’ como es el caso del complejo TATA box TFIID (Janin 1999). Por otro lado observamos que la actividad del agua en $FI > 0.5$ M cambia el componente electrostático de la interacción. Por lo tanto consideramos preferible no medir a concentraciones de fosfato de sodio superiores a 300 mM. Decidimos entonces fijar para posteriores análisis el microentorno de medición del complejo E2C-ADN por ITC a 200 mM de fosfato de sodio pH 7.

Bases termodinámicas de la discriminación de secuencia de ADN.

Los resultados mostrados en el Capítulo anterior usando datos tomados de la literatura validan la utilización de la teoría de Berg-von Hippel para calcular las energías correspondientes la unión de todos los sitios de alfapapilomavirus. Con el fin de profundizar la validación de la teoría propuesta, medimos la energía de unión de E2C-16 a siete sitios de ADN incluyendo los cuatro sitios canónicos del genoma HPV-16 (BS1 a BS4) y tres sitios de diferente secuencia mediante ITC (**Figura 4** y **Tabla 2**). Medimos la unión a una variante natural del BS3 presente en un 25% de los subtipos existentes de HPV-16 (146 subtipos analizados) el cual fue llamado BS3*; a un sitio de unión propio del genoma de papilomavirus bovino (Bov) y a un sitio de unión diseñado con un espaciador que minimice la pre-curvatura del ADN, compuesto por las bases ACGT (BSR) (**Figura 5** y **Tabla 2**). La mutación en BS3* corresponde a la inserción de una adenina en la posición 4 del espaciador de BS3 y aparece en un 46% (17 subtipos) en pacientes con cáncer de laringe y en un 41 % (15 subtipos) en pacientes con cáncer cervical de Turquía, mientras que los restantes subtipos encontrados se encuentran en pacientes de Uganda con cáncer de pene. Realizamos todos los experimentos a 25°C en 200 mM de fosfato de sodio pH 7 y 0.2 mM DTT.

Como primera observación, la unión de E2C-16 a las siete secuencias de ADN detalladas en la **Tabla 2**, presentó una estequiometría >0.8 , y la relación entre la constante de disociación y la concentración de E2C-16, o parámetro C, es superior a 8 (**Tabla 2**), validando el ajuste de los datos a un modelo de un sitio de unión para obtener la constante de disociación de la interacción E2C y ADN. Por lo tanto en una misma condición de medición podemos estudiar tanto la jerarquía de afinidades como la disección termodinámica de la unión de E2C-16 a siete sitios.

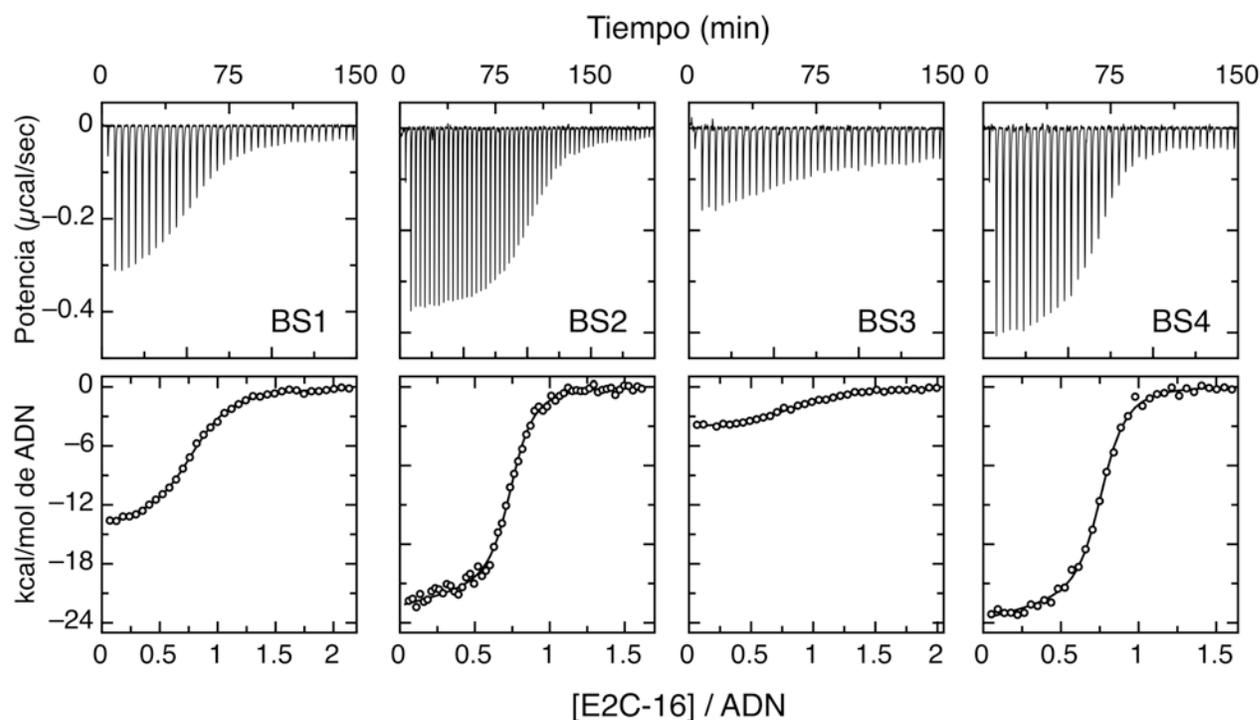


Figura III.4. Jerarquía de unión de E2C-16 a los sitios de ADN canónicos seguida por ITC. Los paneles de arriba muestran los datos de ITC crudos, los paneles de abajo los respectivos datos de ITC integrados y normalizados por concentración, donde la línea sólida corresponde al ajuste a un modelo de un sitio de unión a los datos correspondientes a los complejos entre E2C-16 y BS1, BS2, BS3 y BS4. Tampón químico de medición: 200 mM fosfato de sodio pH 7, 0.2 mM DTT a 25°C.

Estudios biofísicos y bioinformáticos realizados anteriormente en otros laboratorios indican que la alta afinidad de la unión de E2C y ADN se encuentra asociada con la secuencia canónica en los hemisitios ACCG/CGGT y con un espaciador encorvado rico en A/T (Crothers 1998; Hines *et al.* 1998; Falconi *et al.* 2008), mientras que la presencia de un espaciador recto rico en dinucleótidos CG conlleva a uniones más débiles (Barbic *et al.* 2003; Djuranovic *et al.* 2004; Zhang *et al.* 2004). La caracterización termodinámica de la interacción entre E2C-16 y los siete sitios de unión BS1, BS2, BS3, BS4, BS3*, BSR y BSbov (**Tabla 2**) medidos por ITC (**Figuras 4 y 5**) concuerda excelentemente con este conocimiento acumulado sobre la interacción E2C-ADN. En efecto, los sitios BS2 y BS4 poseen hemisitios canónicos y un espaciador rico en A-T y forman los complejos más fuertes con E2C-16 ($\Delta G = -9.6 \pm 0.2$ kcal/mol y -9.8 ± 0.1 kcal, respectivamente). Luego sigue el BS1 con una afinidad de -9.0 ± 0.1 kcal/mol, con hemisitios canónicos pero con una citosina en la posición 4 del espaciador.

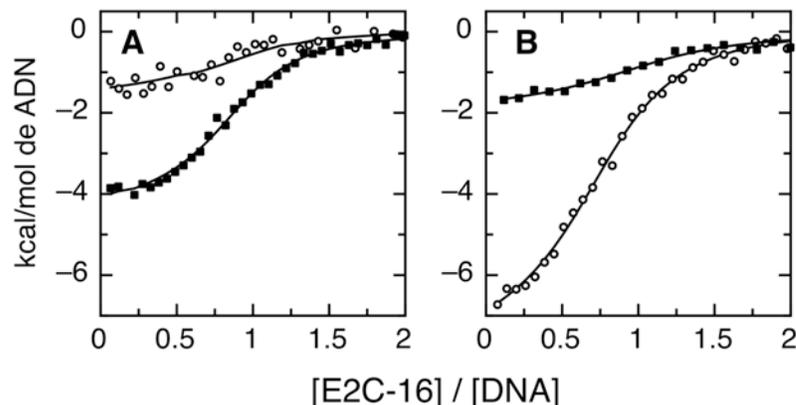


Figura III.5. Reconocimiento por E2C-16 de los sitios de ADN BS3 y BS3* y a los sitios BSR y Bov. Los paneles muestran los datos de ITC integrados y normalizados por concentración ajustados a un modelo de un sitio de unión. (A) Comparación entre la unión de E2C-16 al sitio de ADN BS3 (■) y BS3* (○). (B) Comparación entre la unión de E2C-16 al sitio de ADN BSR (■) y Bov (○). Tampón químico de medición: 200 mM fosfato de sodio pH 7, 0.2 mM DTT a 25°C.

Los hemisitios de BS3 y su variante natural BS3* difieren en una posición entre si y de cuatro de la secuencia canónica. Llamativamente sus afinidades disminuyen en forma similar (ΔG de -8.4 kcal/mol y -8.5 kcal/mol respectivamente) a pesar de tener el segundo hemisitio diferente. Finalmente, el sitio BSR y el sitio Bov forman los complejos más débiles a -8.2 kcal/mol, debido a la presencia del dinucleótifo CG en el espaciador. Nuestras medidas de ITC permiten por primera vez disectar el origen termodinámico de la discriminación de secuencia en componentes entálpicos y entrópicos.

La energía libre de unión ΔG para los complejos más fuertes (BS1, BS2 y BS4) se originan mayoritariamente en la compensación de cambios en ΔH y en $-T\cdot\Delta S$, siendo la entalpía el mayor componente favorable y en la entropía el componente más desfavorable (**Figura 6A** y **Tabla 2**). Existe una compensación isotérmica entrópico-entálpica en toda esta serie de sitios de unión de E2, como podemos apreciar en la correlación lineal de ΔH con ambos $-T\cdot\Delta S$ y ΔG (**Figura 6B** y **6C**, respectivamente). La presencia de ambas correlaciones, la independencia de la medición de ΔH y de ΔG en los ensayos de ITC y el bajo error promedio de 0.4 kcal/mol en la determinación de ΔH comparada con el rango de 1.6 kcal/mol en ΔG observado (de -8.3 a -9.6 kcal/mol) aseguran que las correlaciones reflejan un fenómeno físico y no un artefacto estadístico (Leffler *et al.* 1963; Jen-Jacobson *et al.* 2000). La presencia de compensación entrópico-entálpica en esta serie de sitios de unión de E2 indica que el modo de unión está conservado para todos los sitios ensayados, a pesar de los cambios extensivos en la secuencia de

ADN. Es decir, los factores que modulan la discriminación de secuencia por E2C son los mismos en todos los complejos.

Tabla III.2. Unión específica de E2C-16 y ADN medida por ITC en 200 mM de fosfato de sodio, pH 7.0, 0.2 mM DTT a 25°C.

Sitio y Secuencia diana [#]	ΔH (kcal/mol)	K_D (nM)	ΔG (kcal/mol)	$-T \cdot \Delta S$ (kcal/mol)	N	C
BS1: AACCGAAACCGGTT	-15.2±0.4	248±29	-9.01±0.07	6.2±0.3	0.8±0.1	26±5
BS2: AACCGAAATCGGTT	-22.7±0.4	90±40	-9.61±0.24	13.1±0.2	0.8±0.1	83±24
BS3: AACCCAAAACCGGTT	-4.8±0.8	764±427	-8.35±0.31	-3.6±1.1	0.9±0.1	13±2
BS3*: AACCCAAAATCGGT	-1.5±0.3	597±196	-8.49±0.19	-7.0±0.5	0.9±0.1	11±7
BS4: AACCGAATTCGGTT	-24.1±0.7	67±2	-9.78±0.02	14.3±0.8	0.8±0.1	75±3
BSR: AACCGACGTCGGTT	-2.2±0.1	1070±235	-8.15±0.15	-6.0±0.2	0.9±0.1	9±2
Bov: GACCGACGTCGGTC	-7.7±0.4	978±74	-8.20±0.05	-0.9±0.1	0.9±0.1	8±2

[#] Solo los nucleótidos que se saben que afectan la afinidad están mostrados, la secuencia original incluye además dos nucleótidos flanqueantes a cada lado de la secuencia mostrada. Todas las secuencias están mostradas en la dirección de 5' a 3' salvo BS3 y el BS3*, para los cuales se muestra la secuencia reversa complementaria lo que facilita una comparación con los otros sitios. El ajuste a un modelo de unión 1:1 de los datos crudos proporciona el cambio en entalpía y la constante de disociación al darse la unión. ΔG es calculado como $RT \cdot \ln(K_D)$, $-T \cdot \Delta S = \Delta G - \Delta H$. N es la estequiometría de la reacción y C es la relación entre la concentración de ligando y la K_D (Wiseman *et al.* 1989). Los parámetros representados provienen de por lo menos dos mediciones ajustadas globalmente (Houtman *et al.* 2007), los errores corresponden al desvío estándar de los ajustes individuales.

Finalmente, como anticipamos en el Capítulo anterior, calculamos la energía de unión para cada uno de los siete sitios de acuerdo a su secuencia nucleotídica (**Tabla 2**) y al alineamiento de todos los sitios del *genus* alfa (**Figura II.3**), ver métodos. En la **Figura 6D** comparamos la energía de unión medida en ITC (**Tabla 2**) con la energía de unión predicha. Notablemente, aún considerando un sitio de unión del genoma bovino, que no pertenece al *genus* alfa, y un sitio de unión diseñado (BSR), los datos correlaciones con un grado de confianza de 0.86 (**Figura 6D**). La coincidencia sorprendente entre los experimentos y la teoría implican que los sitios de unión de E2C en el genoma de papilomavirus evolucionan como se postuló en el Capítulo II y que la teoría puede ser utilizada para predecir la afinidad de los dominios de E2C de alfapilomavirus para cualquier secuencia.

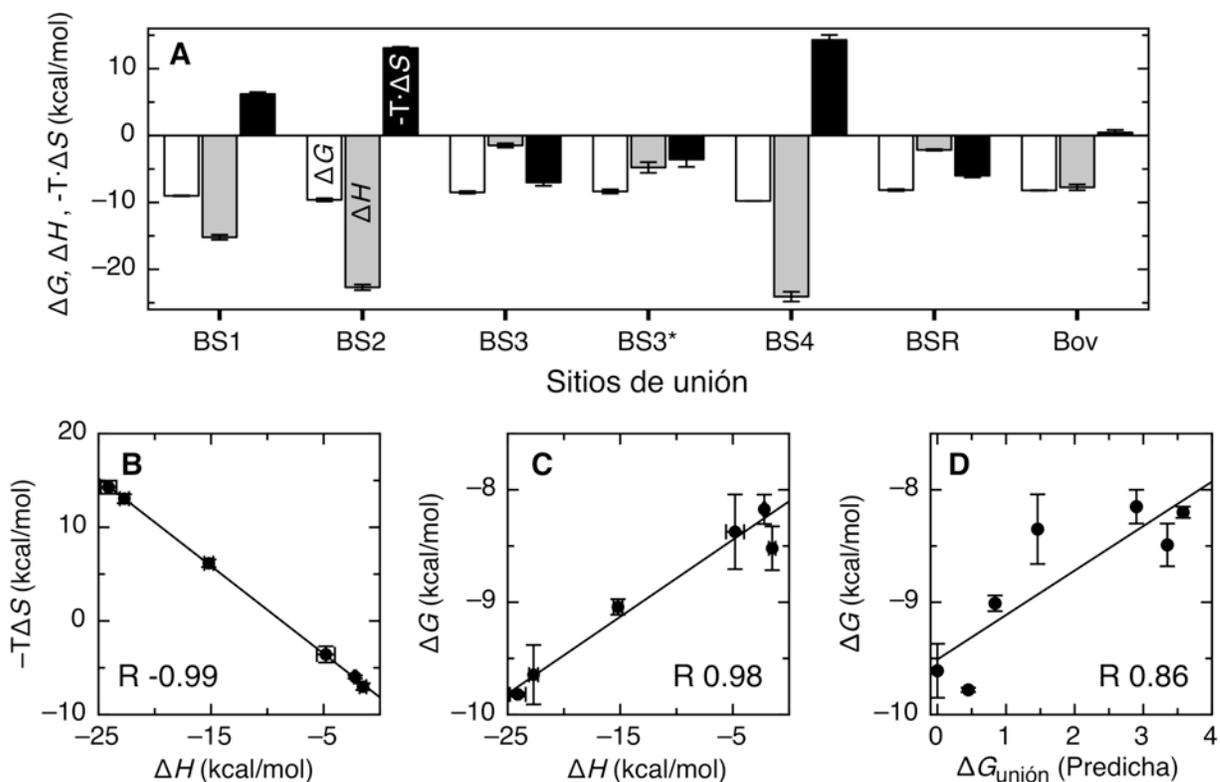


Figura III.6. Disección de la discriminación de secuencia de E2C-16, compensación entálpica-entrópica. (A) Disección termodinámica de la unión de E2C-16 a ADN seguida por ITC. (B) Correlación entre ΔH y $-T\Delta S$. (C) Correlación entre ΔH y ΔG . (D) Correlación entre energías libres predichas y observadas de los complejos E2C-16 y ADN. La energía de unión a la secuencia consenso fue arbitrariamente llevada a cero. Las energías libres predichas se calcularon considerando todas las posiciones nucleotídicas de los sitios. Las energías predichas sin incluir las posiciones 4-5 del sitio de unión, empeoran la correlación con un $R = 0.73$ (no mostrado). Las unidades en todos los casos corresponden a kcal/mol.

La variación en la secuencia de los siete sitios que medimos es considerable, incluyendo cambios en las 14 bases en hasta 9 posiciones conservadas del sitio blanco de E2 (Tabla 2). A pesar de esta variabilidad, la presencia de compensación entalpía-entropía indica que el mecanismo de unión es el mismo en todos los complejos. Por lo tanto, la robustez del dominio E2C-16 para formar complejos específicos aún con variaciones sustanciales en la secuencia del sitio blanco, parece ser un requerimiento fisiológico en papilomavirus permitiendo obtener dentro del mismo modo de unión una jerarquía definida en los cuatro sitios presentes en el genoma. Resulta interesante que E2C-16 reconoce prácticamente con la misma afinidad a los sitios altamente divergentes BS3 y BS3*. En el Capítulo siguiente ahondaremos en el estudio de discriminación de secuencia variando drásticamente la secuencia de E2C, es decir, utilizando 5 proteínas E2C homólogas con el objetivo de analizar la plasticidad del dominio dimérico al ejecutar su función primordial.

Lectura indirecta del lado de la proteína, dependencia del pH.

Como mencionamos desde el comienzo del Capítulo II, la propiedad molecular de la afinidad jerárquica se basa principalmente en la lectura indirecta o pre-curvatura del ADN de unión de E2. Esta curvatura se encuentra definida por las bases espaciadoras entre los hemisitios de secuencia ACCG, que no forman contactos directos con la proteína. A continuación, evaluaremos la presencia de otros fenómenos de “lectura indirecta” mediante cambios en el pH del microentorno de medición.

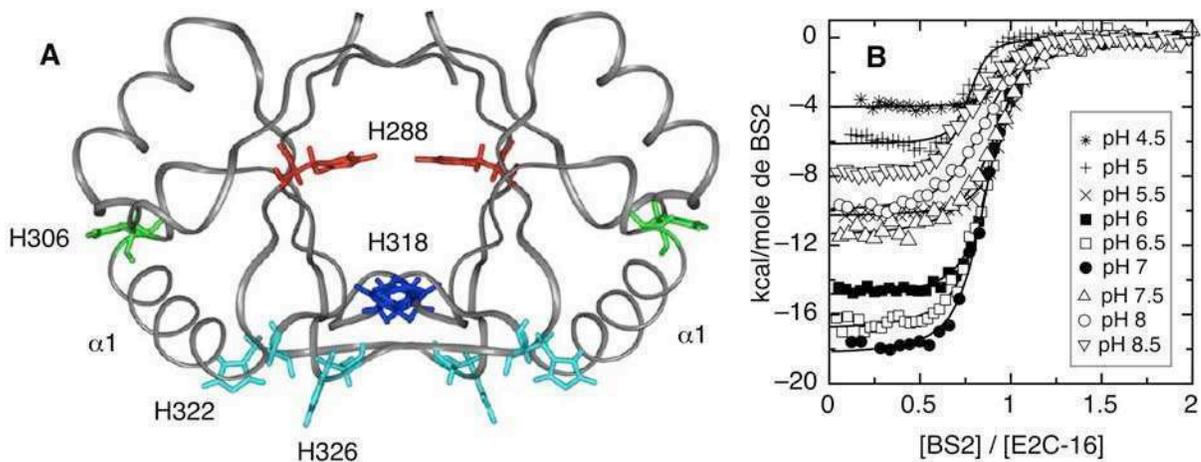


Figura III.7. Unión específica a BS2 de E2C-16 en función del pH en tampón químico triple a 25°C. (A) Entorno molecular de las histidinas en la estructura de E2C-16 (Nadra *et al.* 2004), se destacan las cadenas laterales de las cinco histidinas: H288 (rojo), H306 (verde), H318 (azul) y H322 y 326 (magenta). **(B)** Isothermas de unión resultantes de la integración de los calores específicos en función del tiempo en el rango de pH de 4.5 a 8.5.

El pH como variable puede generar eventos de protonación asociados a procesos de isomerizaciones en residuos sensibles, modificando de esta forma la energía libre para la formación de un complejo. Los residuos de histidina presentan además un equilibrio tautomérico sensible a pequeños cambios en el pH. La arquitectura proteica puede sufrir cambios conformacionales tanto locales como de largo alcance de acuerdo con la protonación y población isomérica de las histidinas. Finalmente, dichos cambios en la tautomerización por protonación de las histidinas pueden ser biológicamente relevantes dado que el pK_a de las histidinas es 6.3, un valor cercano a pHs fisiológicos. Por lo tanto la histidina puede jugar el rol de fino regulador funcional sensible a cambios en la protonación del microentorno. Nuestra proteína en estudio presenta cinco histidinas altamente conservadas (Bose *et al.* 2007; Eliseo *et al.* 2009), cuya disposición espacial se representa en la **Figura 7A**. Exploramos la dependencia de la unión de

E2C-16 y BS2 con la variación del pH por ITC (**Figura 7B**). Los experimentos fueron llevados a cabo a 25°C en un tampón químico de tres componentes que asegura una fuerza iónica de 0.3 M constante en el rango de pH 4.5 a pH 8.5 (Ellis y Morrison 1982). En todas las condiciones de pH medidas, la estequiometría observada resultó ser de 0.8 o mayor y la relación entre la constante de asociación y la concentración de E2C, valor C (Wiseman *et al.* 1989), se encontró por arriba de 40 (**Tabla 3**), validando el ajuste a un modelo de un sitio de unión en todas las condiciones ensayadas y las constantes de disociación obtenidas. A valores de pH más allá del rango abordado, E2C no resultó estable en solución dando lugar a titulaciones isotérmicas con estequiometrías anómalas (datos no mostrados).

Tabla III.3. Unión específica de E2C-16 a BS2 en función del pH.

pH	ΔH (kcal/mol)	K_D (nM)	ΔG (kcal/mol)	$-T \cdot \Delta S$ (kcal/mol)	N	C
4.5	-4.1±0.1	26±4	-10.40±0.09	-6.4±0.1	0.9±0.1	360
5.0	-6.2±0.1	35±9	-10.21±0.15	-4.1±0.2	0.8±0.1	234
5.5	-10.3±0.1	52±5	-9.98±0.05	0.3±0.1	1.0±0.1	206
6.0	-14.9±0.1	63±5	-9.86±0.04	5.1±0.1	0.9±0.1	141
6.5	-16.9±0.1	76±7	-9.75±0.05	7.1±0.1	0.9±0.1	112
7.0	-18.3±0.1	78±5	-9.64±0.04	8.7±0.1	0.8±0.1	110
7.5	-11.5±0.1	111±11	-9.52±0.06	2.0±0.1	0.9±0.1	75
8.0	-10.0±0.1	139±12	-9.39±0.05	0.6±0.1	0.8±0.1	61
8.5	-8.2±0.1	165±16	-9.29±0.06	-1.1±0.1	0.8±0.1	46

La fuerza iónica total es 0.3 M, independiente del pH. El ajuste a un modelo de unión de un sitio de los datos crudos de ITC proporciona el cambio en entalpía y la constante de disociación al darse la unión. ΔG es calculado como $RT \cdot \ln(K_D)$, $-T \cdot \Delta S$ se calcula como ΔG menos ΔH . N es la estequiometría de la reacción y C es la relación entre la concentración de ligando y la constante de disociación (Wiseman *et al.* 1989). Los errores de ajuste de la entalpía, la constante de disociación y la estequiometría fueron propagados para calcular los restantes.

Si la reacción de unión de E2C a BS2 se encuentra acoplada a otra reacción de entalpía significativa, deberíamos observar una relación $-T\Delta S / \Delta H$ y $\Delta G / \Delta H$ elíptica alrededor de un punto, en lugar de lineal como observamos en la **Figura 6B** y **6C** (Eftink *et al.* 1983). En la **Figura 8A**, graficamos la disección termodinámica correspondiente a la unión de E2C a distintos pHs. Mientras que el ΔG varía apenas 1.1 kcal/mol, el ΔH y el $-T\Delta S$ varían ~14 kcal/mol.

Podemos apreciar que tanto $-T\Delta S$ como ΔH alcanzan valores extremos a pH 7 (**Tabla 3**). Esto origina la forma elíptica en la gráfica de el ΔH versus $-T\Delta S$ (**Figura 8B**), y de el ΔG versus ΔH (**Figura 8C**), indicando que la reacción de unión estaría acoplada a otra reacción (Eftink *et al.* 1983). Dado que la concentración de protones en la solución es nuestra única variable experimental en el rango de pH explorado, la unión estaría asociada a al menos un evento de protonación en E2C.

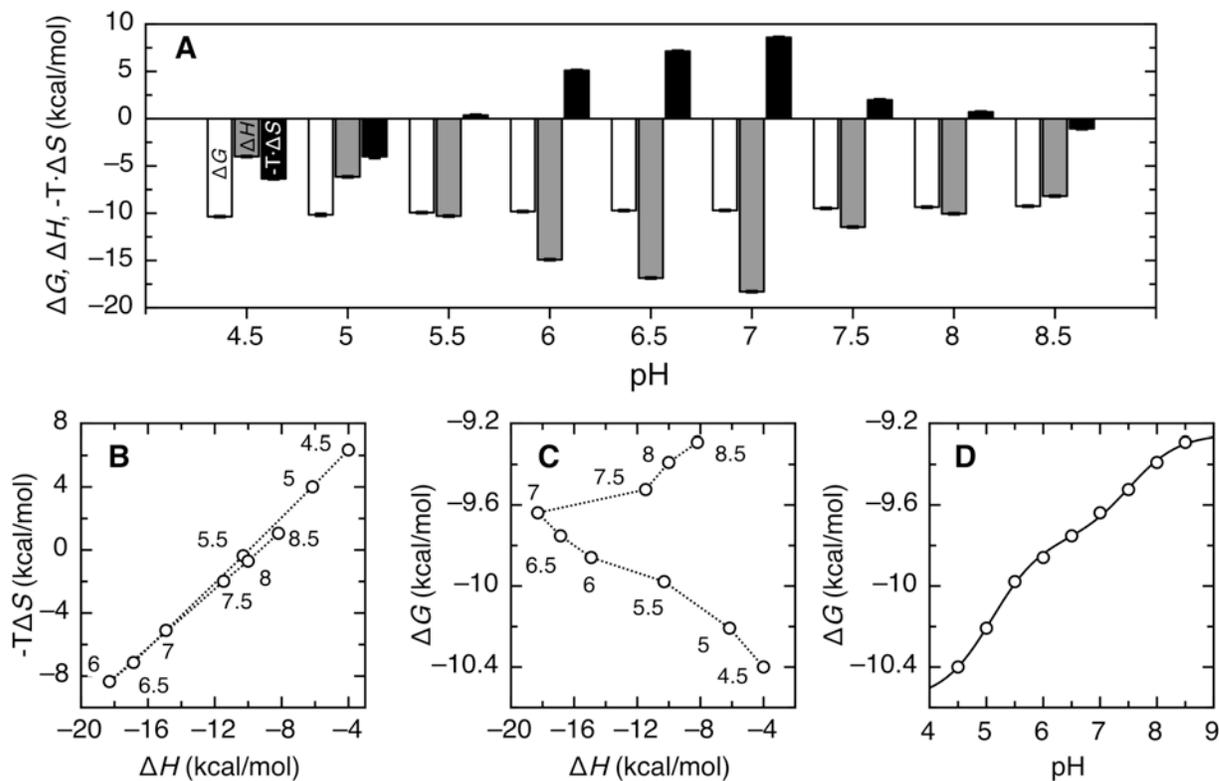


Figura III.8. Disección termodinámica de la unión específica a BS2 de E2C-16 en función del pH. (A) Disección termodinámica de las isotermas de unión para cada condición de pH: ΔG barras vacías, ΔH barras grises y $-T\Delta S$ barras negras. (B) Gráfico de $-T\Delta S$ versus ΔH , se ilustra el pH para cada par de puntos. (C) Gráfico de ΔG versus ΔH , se ilustra el pH para cada par de puntos. (D) Energía libre de unión en función del pH. La línea negra corresponde al ajuste a un modelo de dos residuos titulantes produciendo valores de $pK_a = 4.95 \pm 0.08$ y 7.42 ± 0.09 en E2C libre y valores de $pK_a = 5.24 \pm 0.07$ y 7.61 ± 0.09 en E2C unida a BS2. Composición del tampón triple de medición: 100 mM Tris, 50 mM MES, 50 mM acetato de sodio NaCl y 0.2 mM DTT.

En efecto, al analizar la energía libre para la unión específica E2C-BS2 en función del pH, encontramos dos transiciones (**Figura 8D**). Esta observación indica que por lo menos dos grupos funcionales cambian su pK_a al formarse el complejo (Baker y Murphy 1996). Notablemente, la unión de E2C a BS2 es más afín a pHs ácidos que básicos (**Figura 8D** y **Tabla 3**) y para ambas transiciones, mostrando que el pK_a de ambos grupos es mayor en el complejo

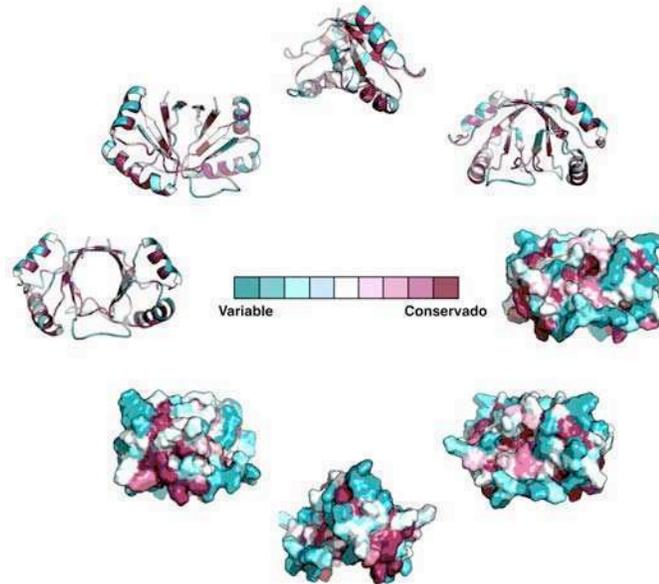
que en los reactivos libres (Baker y Murphy 1996). Ajustamos los datos (**Figura 8D**) a un modelo de dos residuos titulantes independientes con el fin de obtener los valores de pK_a relevantes (detallado en métodos) (Baker y Murphy 1996). Uno de los grupos titulantes que influye en la unión a BS2 presenta un pK_a de 4.95 ± 0.07 en el estado libre y un pK_a de 5.24 ± 0.07 en el complejo, mientras que el otro grupo titulante presenta un pK_a de 7.42 ± 0.09 en el estado libre y un pK_a de 7.61 ± 0.09 en el complejo. Esto implica que existen por lo menos dos residuos en E2C responsables de la regulación de la afinidad por BS2 al variar la acidez del medio.

Los residuos responsables de las transiciones observadas en ITC fueron determinados por análisis estructurales de E2C en función del pH medidos por CD y RMN llevados a cabo en nuestro laboratorio y gracias a la colaboración con el laboratorio del Dr. Daniel Cicero, de la universidad Tor Vergara de Roma, Italia (Eliseo *et al.* 2009). Se logró identificar en detalle el entorno molecular de los residuos de histidina de E2C, tanto en el estado libre como en el complejo con BS2 (Eliseo *et al.* 2009). Se identificó que el acoplamiento entre la protonación y la unión a BS2 observado en ITC con un pK_a de alrededor de 7 correspondió a la protonación de la histidina 318 (**Figura 7A**, en azul), mientras que el evento de pK_a ácido observado en ITC correspondió a la histidina 288 (**Figura 7A**, en rojo). Sorprendentemente, el vínculo entre acidificación del entorno y el aumento de afinidad no estuvo relacionado la lectura directa, es decir con contactos directos al ADN, sino con cambios cooperativos en dinámica de E2C. Por un lado, la H288 es un residuo que se encuentra enterrado en el centro hidrofóbico de E2C, lejos de la superficie de contacto con ADN. Por otro lado, los cambios en el pH que repercuten en la protonación de la H318 generan principalmente efectos de largo alcance influyendo en la dinámica del barril beta que da la identidad de dímero a E2C (Eliseo *et al.* 2009). Desde el punto de vista de la estabilidad, al aumentar de pH 5.6 a pH 8, la constante de disociación de desnaturalización al equilibrio se incrementa unas 200 veces (Mok *et al.* 1996). E2C presenta entonces una conexión sutil entre la termodinámica de plegamiento y la dinámica conformacional permitiendo modular finamente su función de unión a ADN. La alta conservación de los residuos de histidina involucrados (Bose *et al.* 2007; Eliseo *et al.* 2009), indica que esta “lectura indirecta del lado de la proteína” es clave para la dinámica o “respiración, transpiración” del dímero E2C y su función como factor de transcripción y replicación.

Hemos visto entonces que en la lectura de una secuencia de ADN por E2C actúan varios mecanismos en forma simultánea. A continuación estudiaremos como afecta la divergencia de secuencia en E2C mediante el uso de cinco dominios homólogos naturales y un análisis intensivo de estabilidad y función.

Capítulo IV

“Plegamiento, estabilidad conformacional y reconocimiento de ADN de 5 dominios de E2C correspondientes a los tipos de papilomavirus humano de mayor relevancia”



Le feu et la chaleur fournissent des moyens d'explication dans les domaines les plus variés parce qu'ils sont pour nous l'occasion de souvenirs impérissables, d'expériences personnelles simples et décisives.

Le feu est ainsi un phénomène privilégié qui peut tout expliquer.

Gaston Bachelard, La Psychanalyse du Feu

Nos propusimos purificar cinco dominios homólogos de la proteína E2C, con el fin de estudiar el rol de la variabilidad evolutiva en su estabilidad y la funcionalidad.

Contexto evolutivo y estructura primaria de cinco dominios E2C recombinantes.

La longitud aminoacídica de cada dominio fue establecida en base a tres consideraciones, (1) la longitud natural del dominio E2C generado por empalme alternativo (*splicing* en inglés) del marco de lectura abierto de E2 (Zheng y Baker 2006), (2) los estudios estructurales y funcionales previos de diferentes dominios E2C: E2C-06 (Dell *et al.* 2003; Hooley *et al.* 2006), E2C-16 (Hegde y Androphy 1998; Nadra *et al.* 2004), E2C-18 (Kim *et al.* 2000) y E2C-31 (Liang *et al.* 1996; Bussiere *et al.* 1998) y (3) la conservación de secuencia aminoacídica de E2 del *genus* alfafapilomavirus (**Figura 1**), como analizaremos a continuación.

Elegimos representar el alineamiento múltiple de secuencias de E2 del *genus* alfafapilomavirus con el visualizador *LogoBar* (Perez-Bercoff *et al.* 2006; Larkin *et al.* 2007) (**Figura 1A** y **1B**). El *LogoBar* como representación gráfica es muy similar al *logo de secuencia* visto en el Capítulo anterior (Schneider y Stephens 1990), 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 (Schneider y Stephens 1990; Crooks *et al.* 2004; Perez-Bercoff *et al.* 2006; Larkin *et al.* 2007). 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 (Perez-Bercoff *et al.* 2006; Larkin *et al.* 2007). En la **Figura 1A** observamos que el perfil del alineamiento de secuencias de E2 es consistente con su naturaleza modular. La región con mayor número de sustituciones y mayor presencia de brechas coincide con la región flexible de E2. Por el contrario, las regiones de menor variabilidad y menor presencia de brechas coinciden con los dominios funcionales de la proteína, el dominio N-terminal de transactivación y el dominio C-terminal de dimerización y unión a ADN. Por lo tanto si podemos identificar visualmente el límite de los dominios funcionales de E2, de esta manera es posible extrapolar la longitud de cada módulo para todas las secuencias de E2 de alfafapilomavirus.

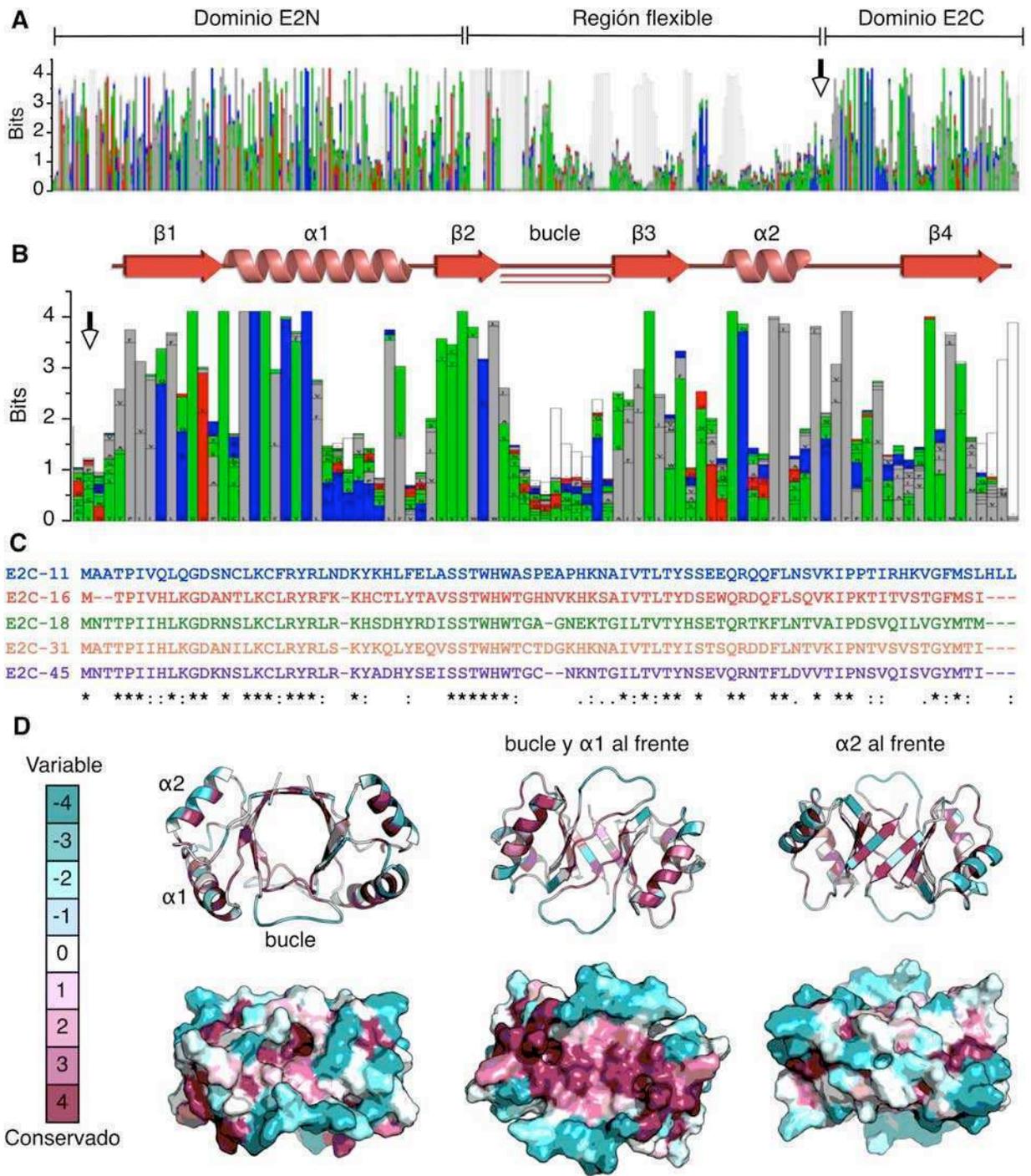


Figura IV.1. Variabilidad evolutiva de E2 de alfapapilomavirus. Alineamiento múltiple de secuencias de (A) E2 y (B) del dominio E2C del *genus* alfa, realizados con *Clustal X* (Larkin *et al.* 2007) y visualizados con *LogoBar* (Perez-Bercoff *et al.* 2006). Los residuos están coloreados en base a su naturaleza química: azul básicos, rojo ácidos, verde polares, gris hidrofóbicos y en blanco las brechas (*gaps*). La estructura secundaria extendida corresponde al dominio E2C-16 (PDB: 1zzf), generada en (Laskowski 2009). La flecha indica el extremo N-terminal del dominio E2C. (C) Alineamiento de los dominios E2C en estudio. Siendo ‘*’ residuos idénticos, ‘:’ sustituciones conservadas (bioquímica) y ‘.’ sustituciones semi-conservadas (tamaño). (D) Arquitectura evolutiva del dominio E2C generada a partir del alineamiento del *genus* alfa y de la estructura de E2C-16 (PDB: 1r8p, (Nadra *et al.* 2004)) con *Consurf* (Ashkenazy *et al.* 2010) y *Pymol* (DeLano). Los residuos conservados se muestran en magenta y los variables en turquesa en la estructura secundaria (arriba) y la superficie expuesta al solvente (abajo).

En la **Figura 1B** mostramos la variabilidad evolutiva del dominio E2C. El extremo N-terminal de E2C, marcado con una flecha en la **Figura 1A y 1B**, coincide con un aumento en la variabilidad de secuencia indicativo de la frontera con la región flexible de E2. Como referencia, incorporamos al perfil del alineamiento la estructura secundaria extendida de E2C-16. Podemos identificar patrones de conservación correspondientes a regiones funcionales y coincidentes con elementos de estructura secundaria. Por ejemplo, la conservación de los de residuos de la hélice $\alpha 1$ de unión a ADN analizada en la **Figura II.1** y la mayoría de las láminas de hoja β . Sin embargo, también abundan regiones altamente variables como es la zona del bucle $\beta 2$ - $\beta 3$ y en la hélice $\alpha 2$.

En conjunto, las tres consideraciones mencionadas arriba nos guiaron para definir el extremo N-terminal de los cinco dominios E2C-11, E2C-16, E2C-18, E2C-31 y E2C-45 cuya secuencia completa se detalla alineada en la **Figura 1C**. Los cinco dominios fueron elegidos por pertenecer a los tipos de HPV con mayor relevancia epidemiológica (Munoz 2000; Schiffman *et al.* 2003; Sichero *et al.* 2006; WHO 2007). Una forma adicional de plasmar la variabilidad a partir del alineamiento entre las cinco homólogas elegidas (**Figura 1C**), consiste en calcular la identidad de secuencia aminoacídica entre todos los pares posibles (Hall 2005):

	E2C-11	E2C-16	E2C-18	E2C-31	E2C-45
E2C-11	<i>ID</i>	55%	46%	57%	45%
E2C-16		<i>ID</i>	52%	66%	52%
E2C-18			<i>ID</i>	59%	77%
E2C-31				<i>ID</i>	60%
E2C-45					<i>ID</i>

El grupo de las cinco proteínas (**Figura 1C**) genera 10 pares de dominios distintos con variaciones en la identidad de secuencia proteica desde 45% para el par E2C-45/E2C-11 hasta 77% para el par E2C-45/E2C-18, con un promedio de $56 \pm 7\%$. Sumado a que de los 84 residuos en promedio que conforman las cinco proteínas (**Tabla 1**), solo 35 son idénticos entre ellas (**Figura 1C**, “*”), resulta evidente que la variación de secuencia en el grupo de las cinco E2C homólogas es extensa. Los 73 dominios E2C del todo *genus* alfapapilomavirus generan 2628 pares de homólogas distintos cuya identidad de secuencia promedio calculada a partir del alineamiento de la **Figura 1B**, es de $48 \pm 10\%$. Comparando este último promedio con el promedio de identidad del grupo de las cinco E2C, notamos que dentro del desvío estándar, la

identidad promedio del *genus* alfa incluye la identidad promedio del grupo conformado por cinco dominios. Esta observación sugiere que la extensa divergencia de secuencia presente en el grupo de los cinco dominios es cuantitativamente representativa de la divergencia de secuencia de todos los dominios E2C del *genus* alfa.

Por otro lado, analizamos cualitativamente si la divergencia de secuencia del grupo elegido es representativa de la existente en alfapapilomavirus. Con este fin, representamos las posiciones residuales del alineamiento de E2C del *genus* alfa (**Figura 1B**) de forma tal que concuerden con las posiciones residuales de los cinco dominios E2C (**Figura 1C**). A grandes rasgos, la posición de residuos invariables (“*”) corresponde con la altura máxima de barras del *LogoBar*, la ubicación de los residuos con propiedades bioquímicas (“:”) y estéricas (“.”) semejantes corresponde con alturas medias de barras del *LogoBar*, la posición los residuos muy diferentes (“ ”) coinciden con las barras de mayor variabilidad y finalmente la ubicación de las brechas (“-”) coincide con la presencia de barras blancas en el *LogoBar*. Podemos entonces concluir que tanto las zonas conservadas como las variables y hasta las que presentan brechas se ven reflejadas en ambos alineamientos a pesar que el de la **Figura 1C** corresponde a solo cinco secuencias. Concluimos que el grupo de cinco dominios es representativo de la variación evolutiva que se ve al analizar todos los dominios de alfapapilomavirus, siendo entonces un excelente muestreo de variabilidad de secuencia dentro del *genus*.

En la **Figura 1D** representamos la divergencia de E2C de alfapapilomavirus en la estructura tridimensional del dominio E2C-16 (PDB: 1r8p, (Nadra *et al.* 2004)) con el fin de visualizar la variabilidad en las regiones estructurales que se exponen al solvente. La región interna de la molécula forma un núcleo hidrofóbico conservado que no se expone al solvente (ver color magenta en las láminas β , **Figura 1D**) que indica una fuerte presión evolutiva en la región de dimerización de E2C, siendo ésta una conservación que “mira hacia adentro” de la molécula. Por otra parte, la zona más conservada de la región que mira “hacia afuera” de la molécula comprende principalmente la región de unión a ADN (ver color magenta **Figura 1D**, representación del medio) como describimos en el Capítulo II. En el resto de la superficie expuesta de E2C predomina la variabilidad de secuencia (ver color turquesa **Figura 1D**), incluyendo tanto regiones flexibles, por ejemplo el bucle situado entre las láminas β 2- β 3, como elementos de estructura secundaria, como se puede apreciar para la hélice α 2. Por lo tanto, la variabilidad entre los cinco dominios es extensa y se localiza principalmente en la superficie

expuesta al solvente. El análisis de variabilidad evolutiva presentado nos contextualiza arquitecturalmente el sistema en estudio y nos permitirá racionalizar en forma relativa los resultados de los ensayos que abordaremos en el presente y en el siguiente Capítulo. Concluimos que aún siendo una muestra pequeña de solo cinco de los 73 tipos virales pertenecientes a 3 de las 14 especies del *genus* alfacapilomavirus (**Figura 1 Introducción**) (Büchen-Osmond 2003; de Villiers *et al.* 2004; Bernard *et al.* 2010), el grupo representa aceptablemente la variabilidad de la secuencia de E2C del *genus* alfa, permitiéndonos el atrevimiento de extrapolar los resultados que obtengamos a todas las E2C del *genus* alfa (ver Discusión).

Tabla IV.1. Características de la estructura primaria de las proteínas recombinantes.

Proteína	Número de residuos.	pI^b	$\epsilon_{280\text{ nm}}^b$ ($\text{mM}^{-1}\cdot\text{cm}^{-1}$)	Masa teórica ^b (Da)	Masa medida ^a (Da)
E2C-11	87	9.34	15.47	10069.6	10033.2 ± 27.1
E2C-16	81	9.80	20.97	9382.9	9383.2 ± 31.0
E2C-18	82	9.73	16.96	9505.8	9514.2 ± 12.5
E2C-31	83	9.43	18.45	9510.9	9510.7 ± 16.0
E2C-45	81	9.30	18.45	9337.6	9335.1 ± 21.7

^a: Medido en MALDI-TOF, la desviación estándar corresponde a la resolución de la medida calculada como ancho completo situado en la mitad del máximo (FWHM). ^b: Calculado a partir de las secuencias mediante la aplicación en línea ProtParam (Gasteiger *et al.* 2005).

Las cinco proteínas homólogas recombinantes fueron purificadas de acuerdo al protocolo de purificación de referencia puesto a punto con el dominio E2C de HPV16 (Mok *et al.* 1996). Las diferencias entre los cinco dominios en el principal paso de purificación, la unión electrostática e inespecífica a la heparina, resultaron mínimas (ver métodos), en buen acuerdo con las leves variaciones en el punto isoeléctrico de los dominios (pI , **Tabla 1**). El grado de pureza para los cinco dominios homólogos resultó ser >95% de acuerdo al gel de poliacrilamida al 15 % (**Figura 2, primera fila, recuadro**). Garantizando la ausencia de modificaciones en la estructura primaria de los dominios E2C-11, E2C-16, E2C18, E2C-31 y E2C-45, el peso molecular teórico de cada monómero correspondió dentro del error de medición con peso molecular medido para las cinco proteínas por espectrometría de masas MALDI-TOF (**Tabla 1 y Figura 2, primera fila**).

Análisis espectroscópico de la estructura secundaria, terciaria y cuaternaria.

Habiendo corroborado la estructura primaria de las proteínas homólogas, analizamos a continuación los tres niveles estructurales restantes de los cinco dominios en solución, es decir su conformación espacial utilizando herramientas espectroscópicas (**Figura 2**).

La señal de CD refleja un promedio de las poblaciones de estructura secundaria presentes en la molécula. Podemos apreciar que la forma espectral y la intensidad de la señal varían considerablemente entre los cinco dominios (**Figura 2, segunda fila**). Al someter a los dominios a condiciones desnaturalizantes (**Figura 2, segunda fila, línea punteada**), la señal cambia notoriamente, correspondiendo a un espectro característico de falta de estructura (Gray 1996), evidenciando la presencia de estructura secundaria en el estado plegado. Se ha visto anteriormente que mutaciones sitio dirigidas simples y dobles en todos los residuos de la hélice $\alpha 1$ de proteína E2C-16, es decir la Introducción de cambios mínimos en la identidad de secuencia aminoacídica (en 97-98%) de E2C, no altera la forma y/o intensidad de la señal del espectro de CD (Ferreiro *et al.* 2005). Sin embargo ciertas mutaciones puntuales sí mostraron importantes cambios en la señal de CD respecto a la forma nativa, con una señal más intensa y característica de estructura secundaria de tipo beta (Ferreiro 2003). Esto indica que la intensidad y forma de la señal de CD de la proteína E2C puede ser sensible. Por lo tanto, con identidades de secuencia oscilando en el rango de 45% al 77% es esperable obtener amplias diferencias como las observadas experimentalmente entre las cinco proteínas analizadas (**Figura 2, segunda fila**). No obstante, en todos los casos apreciamos dos mínimos en la señal de CD, uno a 210 nm y otro 224 nm. Estos mínimos indican la presencia de un componente de estructura secundaria helicoidal en presencia de estructura secundaria de tipo hoja beta y de otras contribuciones, ya que los mínimos se alejan ligeramente de los usuales para estructura helicoidal de 208 nm y 222 nm, y tampoco se observa un único mínimo en el rango 214-218 nm característico de estructura tipo lámina beta (Gray 1996).

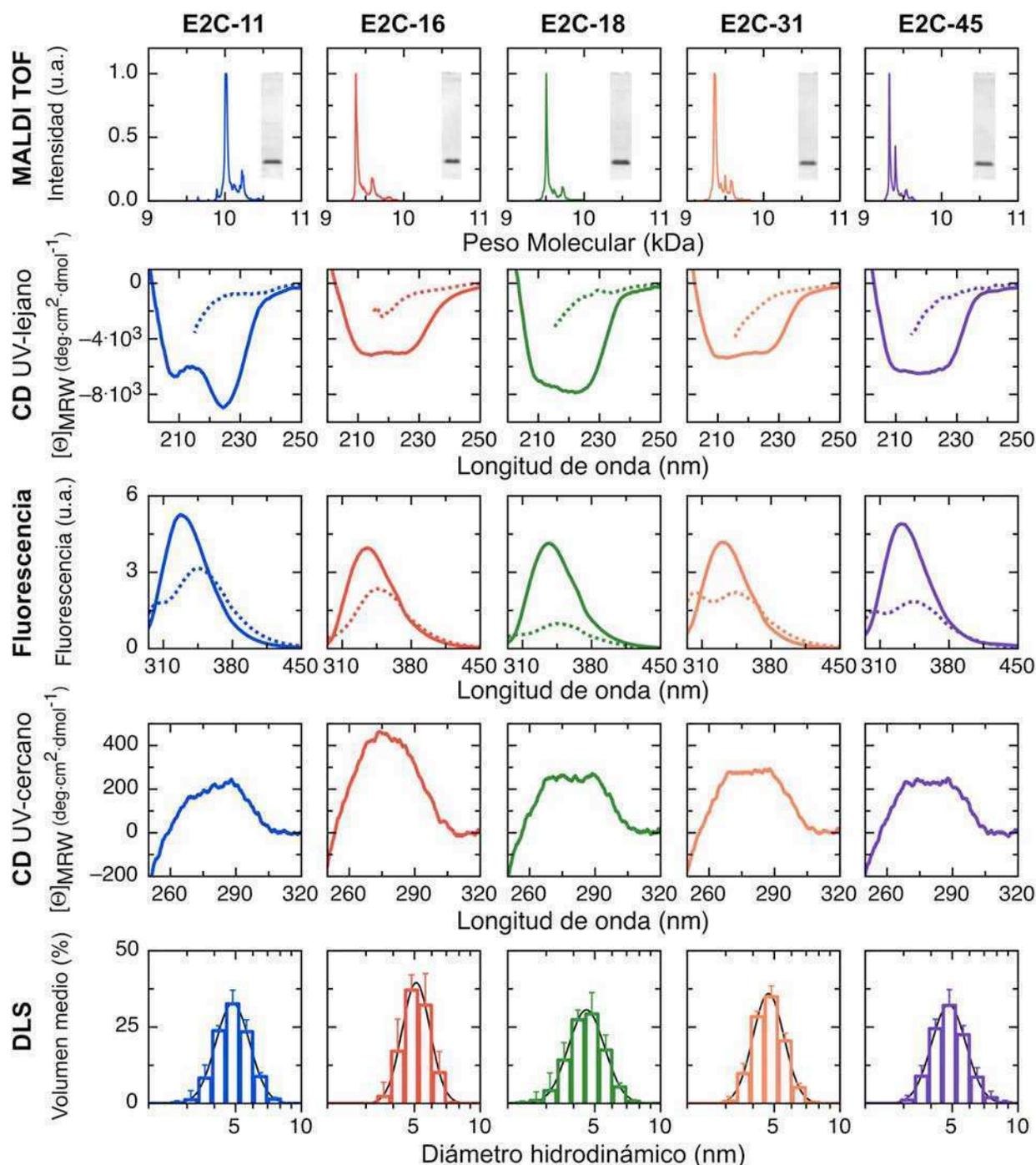


Figura IV.2. Análisis de la estructura primaria, secundaria, terciaria y cuaternaria de E2C-11, E2C-16, E2C-18, E2C-31 y E2C-45 en, monitoreada por Maldi-TOF, CD, fluorescencia y DLS. Primera fila: estructura primaria medida por Maldi-TOF y gel SDS page al 15%. **Segunda fila:** estructura secundaria medida por espectro de CD en el UV lejano en condiciones nativas (línea continua) y desnaturalizantes (línea punteada). **Tercera fila:** estructura terciaria seguida por espectro de fluorescencia de triptófanos en condiciones nativas (línea continua) y desnaturalizantes (línea punteada). **Cuarta fila:** estructura terciaria medida por espectro de dicroísmo circular en el UV cercano. **Quinta fila:** estructura cuaternaria medida por dispersión dinámica de la luz (DLS). En todos los experimentos, la condición nativa corresponde al tampón químico compuesto por 50 mM fosfato de sodio pH 7, 1 mM DTT a 25°C y la condición desnaturalizante corresponden a 50 mM Fosfato pH 7, 1 mM DTT, 3.9 M Gdm.Cl a 25°C. Cada fila fue llevada a la misma escala para facilitar la comparación.

Analizamos la estructura terciaria por fluorescencia de triptófanos y por CD en el UV cercano (**Figura 2, tercera y cuarta fila** respectivamente). La presencia de estructura terciaria consolidada en las cinco proteínas se evidencia como vimos para CD en el UV lejano al comparar la diferencia espectral entre condiciones nativas y desnaturalizantes. El máximo en señal de fluorescencia de triptófanos se encuentra en longitudes de onda menores que el máximo de fluorescencia de triptófanos en condiciones desnaturalizantes, indicando que este residuo se encuentra inmerso en ordenamientos de estructura secundaria formando estructura terciaria en los cinco dominios (Edelhoc 1967). A su vez, la presencia de señal en el CD en el UV cercano indica un entorno asimétrico del residuo triptófano: de no existir un acomodamiento espacial de la estructura secundaria no se debería distinguir señal a estas longitudes de onda (Gray 1996).

Finalmente, analizamos la estructura cuaternaria de E2C-11, E2C-16, E2C-18, E2C-31 y E2C-45 por dispersión dinámica de la luz (DLS) (**Figura 2, última fila**). El último paso de purificación de los dominios involucra un tamiz molecular del cual se puede estimar cualitativamente la masa molecular de las especies. Todos los dominios presentaron una elusión correspondiente a la masa molecular dimérica esperada (ver Discusión). Utilizamos la técnica de DLS, mencionada en el primer Capítulo, con el fin de obtener datos cuantitativos de la estructura cuaternaria de los dominios. El DLS mide el diámetro hidrodinámico de las moléculas y el conocer la estructura cristalográfica y/o de RMN de varios dominios de E2C (Liang *et al.* 1996; Hegde y Androphy 1998; Dell *et al.* 2003; Nadra *et al.* 2004), permite diferenciar el estado multimérico entre los dominios E2C. El diámetro hidrodinámico para los cinco dominios corresponde a 4.9 ± 1.2 nm para E2C-11, 5.1 ± 1.1 nm para E2C-16, 4.6 ± 1.2 nm para E2C-18, 4.7 ± 1.2 nm para E2C-31 y 4.8 ± 1.2 para E2C-45 (**Tabla 2**). La forma cuasi-globular y el ancho promedio medido en su estructura cristalográfica de RMN corresponde a una esfera de aproximadamente 45 ± 5 Å de diámetro. Por lo tanto, podemos concluir que las cinco proteínas homólogas presentan una arquitectura espacial homodimérica, como esperábamos de acuerdo a la literatura. En conjunto, tanto el análisis de pureza y peso molecular de las cadenas polipeptídicas como la disposición conformacional en todos los niveles estructurales analizada permite profundizar la caracterización de los cinco dominios en cuanto a estabilidad y función.

Desnaturalización al equilibrio, estabilidad conformacional de los dominios.

A continuación estudiamos la estabilidad conformacional de los dominios mediante el estudio de desnaturalización química al equilibrio inducida por Gdm.Cl en 50 mM de fosfato de sodio pH 7 a 25°C y a una concentración proteica de 10 μ M (**Figura 3**). Para ello seguimos la señal de CD en el UV lejano reportando el cambio en estructura secundaria (**Figura 3, primeras dos filas**) y la señal de fluorescencia de triptófanos reportando el cambio en estructura terciaria (**Figura 3, tercer y cuarta fila**) en función de la concentración de Gdm.Cl. Para ambas señales mostramos los datos crudos, es decir todas las curvas espectrales, coloreados de acuerdo a una escala de arco iris, correspondiendo los colores fríos a bajas concentraciones de Gdm.Cl y los cálidos a altas concentraciones. Debajo de los datos crudos de la **Figura 3**, graficamos la señal de CD a 225 nm (círculos vacíos) y el centro de masa de los espectros de fluorescencia (círculos llenos) en función de la concentración de Gdm.Cl (ver métodos).

Se puede apreciar que para todos los dominios E2C, a medida que aumenta la concentración del agente desnaturalizante, las señales espectrales reportan el estado desnaturalizado. En efecto, la elipticidad molar a 225 nm en el estado plegado corresponde a valores de alrededor de -5.10^3 a -9.10^3 deg.cm².dmol⁻¹ y disminuye al aumentar la concentración de Gdm.Cl hasta alcanzar un estado de desorden conformacional caracterizado por valores de $[\theta]_{225\text{nm}}$ cercanos a cero (Gray 1996). Este comportamiento se aprecia en los datos crudos en todo el barrido espectral de CD (220 nm a 260 nm). Todas las proteínas son desplegadas a través de una transición aguda, lo que indica un alto grado de cooperatividad en la desnaturalización al analizar la estructura secundaria. Dicha transición aguda ocurre a distintas concentraciones de Gdm.Cl para las distintas proteínas homólogas indicando diferencias en la estabilidad de los estados nativos y desnaturalizados entre las cinco proteínas.

En la **Figura 3, tercer y cuarta fila**, en forma similar a las diferencias en intensidad y forma de la señal de CD (**Figura 2**), observamos que la aspecto y rango de cambios frente a la presencia de Gdm.Cl varía considerablemente entre los cinco dominios. Los espectros de fluorescencia reportan cambios en el entorno de los cromóforos naturales de los dominios E2C (tirosina y triptófanos) al aumentar la concentración de Gdm.Cl. Brevemente, la tirosina solo se ve afectada en cuanto a su intensidad de fluorescencia, mientras que el triptófano sufre tanto cambios en su rendimiento cuántico como en la posición de su máximo de fluorescencia (Lakowicz 1983).

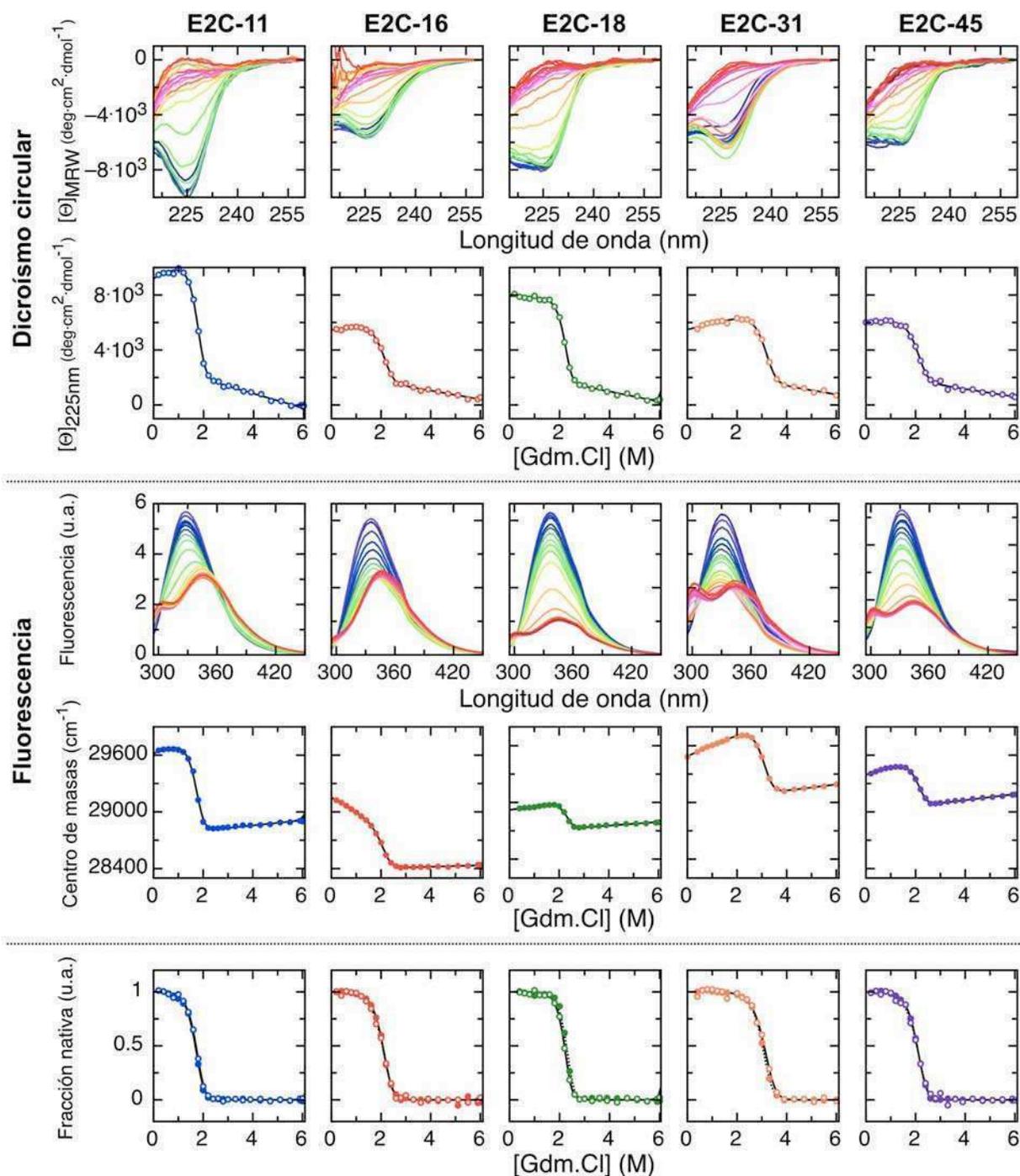


Figura IV.3. Desnaturalización al equilibrio de E2C-11, E2C-16, E2C-18, E2C-31 y E2C-45 en 50 mM fosfato de sodio pH 7, 1 mM DTT a 25°C. Primera y segunda fila: Estructura secundaria monitoreada por espectros de CD en el UV lejano en función de la concentración creciente de Gdm.Cl, coloreados del azul al rojo en escala de arco iris. El modelo de disociación de dos estados fue ajustado (línea negra) a los datos de señal de CD a 225 nm (círculos vacíos). **Tercera y cuarta fila:** Estructura terciaria monitoreada por espectros de fluorescencia de triptófanos en función de la concentración creciente de Gdm.Cl, coloreadas del azul al rojo en escala de arco iris. El modelo de disociación de dos estados (línea negra) fue ajustado a los datos de centro de masa espectral (círculos llenos). La concentración proteica fue de 10 μM . **Quinta fila:** Fracción nativa de las curvas de desnaturalización calculada a partir del ajuste de los datos de CD (círculos vacíos) y fluorescencia de triptófanos (círculos llenos).

Por lo tanto, dado que los cinco dominios varían en número y posición de cromóforos, por ejemplo, E2C-16 es el único de los cinco dominios que posee un triptófano en la α -hélice-2 (**Figura 1B y C**); cada dominio reportará el cambio en estructura terciaria proteica de manera diferente. A pesar de ello, en todos los casos la dependencia de la señal reportera de estructura terciaria con la concentración de Gdm.Cl ocurre con un alto grado de cooperatividad.

De acuerdo a lo descrito en las **cuatro primeras filas** de la **Figura 3**, todos los dominios mostraron transiciones cooperativas en las cuales la estructura secundaria y terciaria se pierden concomitantemente. Esto coincide con resultados previos de desnaturalización al equilibrio medidos para E2C-16 en este (Dellarole *et al.* 2007) y otros tampones químicos (Mok 1996). El modelo de disociación de dos estados para proteínas diméricas ((Mok 1996) y ver métodos) pudo ser ajustado con confianza a todos los datos medidos como es posible reparar en la **segunda y cuarta fila** de la **Figura 3**, líneas negras. En la **quinta fila** de la **Figura 3**, graficamos la fracción nativa calculada a partir del modelo de disociación ajustado a los datos de CD (círculos vacíos) y fluorescencia (círculos llenos). Para las cinco proteínas E2C homólogas, observamos claramente que las fracciones nativas de ambos orígenes se superponen mostrando cooperatividad entre la estructura secundaria y terciaria para la desnaturalización. Consecuentemente, los parámetros de los ajustes de ambas curvas de desnaturalización se pueden analizar en conjunto: en la **Tabla 2** representamos el promedio de los ajustes derivados de las mediciones por ambas señales, cuyo error corresponde al desvío estándar entre los valores.

En primer lugar, el cambio de energía libre entre el estado nativo y el estado desnaturalizado ($\Delta G_{eq(H_2O)}$) medido para las cinco proteínas homólogas presenta valores de rango -14.9 ± 0.1 kcal/mol para E2C-11 a -19.0 ± 0.1 kcal/mol para E2C-18. Este valor termodinámico indica la fuerte naturaleza dímérica de los dominios E2C (**Tabla 2**). En efecto, la constante de disociación K_D , es del orden picomolar o sub-picomolar para las cinco proteínas. Esto garantiza que en la presente tesis los experimentos llevados a cabo con los cinco dominios presentan una población de especies monoméricas despreciable, dado que trabajamos en concentraciones micromolares. Otra manera similar de interpretar la estabilidad termodinámica es analizando el valor de desnaturalización media ($[D_{50\%}]$, **Tabla 2**). E2C-18 y E2C-31 se desnaturalizan a mayores concentraciones de Gdm.Cl que E2C-11, E2C-16 y E2C-45, en acuerdo a un mayor cambio de energía libre ΔG y una menor K_D para estos dos dominios.

Tabla IV.2. Parámetros de la estabilidad conformacional de los dominios E2C.

Proteína	D hidrodinámico ^a (nm)	$\Delta G_{\text{eq(H2O)}}$ ^b (kcal·mol ⁻¹)	K_D ^c (pM)	$[D]_{50\%}$ ^d (M)	m_{eq} ^b (kcal·mol ⁻¹ ·M ⁻¹)
E2C-11	4.9±1.2	-14.86±0.12	12.7±0.9	1.70±0.02	4.72±0.01
E2C-16	5.1±1.1	-15.21±0.69	7.0±2.9	2.07±0.02	4.06±0.30
E2C-18	4.6±1.2	-19.01±0.14	0.011±0.001	2.24±0.06	5.45±0.08
E2C-31	4.7±1.2	-18.58±0.66	0.024±0.009	3.07±0.05	3.84±0.28
E2C-45	4.8±1.2	-15.09±0.76	8.5±3.9	2.06±0.01	4.01±0.35

^a: Diámetro hidrodinámico medido en condiciones nativas por DLS, 50 mM fosfato de sodio pH 7, 1 mM DTT a 25°C. ^b: La diferencia de energía libre entre estado desnaturalizado y plegado, ΔG_{eq} , se relaciona linealmente con la concentración de Gdm.Cl, (Tanford 1968; Pace 1986) según: $\Delta G_{\text{eq}} = \Delta G_{\text{eq(H2O)}} - m_{\text{eq}}[\text{Gdm.Cl}]$, siendo m_{eq} la dependencia del ΔG_{eq} con la concentración de Gdm.Cl y $\Delta G_{\text{eq(H2O)}}$ la diferencia de energía libre entre estado desnaturalizado y plegado en ausencia de Gdm.Cl. ^c: constante de disociación calculada según $\Delta G_{\text{eq(H2O)}} = -RT \cdot \ln K_D$. ^d: desnaturalización media calculada a partir de $[D]_{50\%} = ((RT \cdot \ln [P_t]) + \Delta G_{\text{eq(H2O)}}) / m_{\text{eq}}$, P_t es la concentración proteica total. Los errores de los parámetros termodinámicos corresponden a la desviación estándar del ajuste a un modelo de disociación de dos estados a los datos obtenidos por fluorescencia y CD (ver métodos).

El parámetro m_{eq} de la desnaturalización química, o cooperatividad del proceso, se corresponde empíricamente con el cambio del área expuesta al solvente (ΔASA) al desnaturalizarse la proteína (Myers *et al.* 1995). El parámetro m_{eq} varía considerablemente entre las cinco proteínas (**Tabla 2**) sugiriendo que los dominios presentan diferentes valores de ΔASA en la diferencia entre el estado nativo y el desplegado al equilibrio (Myers *et al.* 1995). Por un lado en la **Figura 1D** apreciamos que los residuos de mayor variabilidad se encuentran ubicados en las zonas expuestas al solvente. Por otro lado, en la **Figura 3, segunda y cuarta fila** observamos la presencia de pendientes en el estado desplegado de diferente magnitud entre los dominios, insinuando la presencia de estructura residual en el estado desplegado, la cual variaría entre los dominios. Por lo tanto, la heterogeneidad observada entre los cinco dominios en el parámetro m_{eq} , puede tener origen ya sea en la superficie expuesta del estado nativo (ASA_N) como en la superficie expuesta del estado desnaturalizado (ASA_D) (Myers *et al.* 1995). Buscamos correlacionar los valores medidos de m_{eq} con diferentes parámetros característicos de ambos estados, con el fin de encontrar una explicación común a los cinco dominios. Sin embargo, no pudimos correlacionar la variabilidad del parámetro m_{eq} entre los dominios ni con datos de la estructura primaria (número de residuos, el peso molecular, pI , **Tabla 1**), ni con datos de la estructura secundaria y terciaria en ambos estados (valores absolutos y relativos de las

señales de CD, pendientes y diferencias espectrales de los estados nativos y desnaturalizados **Figuras 2 y 3**), ni con datos reportando la estructura cuaternaria (radio hidrodinámico, **Tabla 2**, y datos no mostrados). Concluimos que la heterogeneidad en Δ ASA responde a características intrínsecas de cada dominio y no a una característica compartida para el dominio E2C. A continuación profundizamos el análisis de la desnaturalización estudiando el mecanismo de desnaturalización cinético. Globalmente, el análisis de desnaturalización al equilibrio indica que los cinco dominios presentan una elevada estabilidad que garantiza su naturaleza homodimérica en solución y que el proceso de desnaturalización puede describirse simplemente por un modelo de dos estados altamente cooperativo.

Cinéticas de *desnaturalización*.

La desnaturalización cinética de E2C-16 mediante el agregado de agentes químicos es un proceso de dos estados descrito por un decaimiento exponencial (Prat Gay *et al.* 2005; Dellarole *et al.* 2007). A continuación, medimos las cinéticas de la reacción de desnaturalización de los dominios E2C-11, E2C-16, E2C-18, E2C-31 y E2C-45 utilizando de 3 M a 8 M de cloruro de guanidinio como agente desnaturalizante, en las mismas condiciones del estudio al equilibrio, 50 mM fosfato pH 7, 1 mM DTT a 25°C (**Figura 4**). Las dos primeras filas de la **Figura 4** constituyen ejemplos de cinéticas de desnaturalización inducidas por 5 M y 8 M de Gdm.Cl y seguidas por CD y/o fluorescencia. Tanto ambas señales reporteras como distintas concentraciones proteicas presentaron las mismas velocidades de reacción para un dominio dado en una misma condición (datos no mostrados, métodos). Podemos apreciar que para cada dominio, las cinéticas alcanzan un *plateau* a tiempos menores en condiciones de 8 M de Gdm.Cl que en 5 M de Gdm.Cl. Esta observación sugiere una fuerte dependencia de la reacción de desnaturalización con la concentración del agente desnaturalizante. Sin embargo, esto no ocurre para el dominio E2C-45 (**Figura 4, quinta columna**).

Al analizar cuantitativamente las reacciones, en todos los casos salvo para E2C-45, las cinéticas se describen con una función exponencial simple, ($y = A \exp^{-\lambda x}$, siendo λ la constante de velocidad macroscópica), evidenciando la ausencia de intermediarios en la desnaturalización de las homólogas. En el caso de E2C-45 las cinéticas de desnaturalización son descritas por la suma de dos decaimientos exponenciales. En la **Figura 4 tercera fila**, graficamos la dependencia del logaritmo neperiano de λ con la concentración de GdmCl. Para E2C-45, una de las dos

constantes de velocidad (círculos vacíos) es independiente de la concentración de Gdm.Cl. El valor de dichas constantes de velocidad y la independencia respecto de la concentración de Gdm.Cl es característico de isomerizaciones *cis-trans* de enlaces peptídicos Xaa-Pro (Brandts *et al.* 1975). Por lo tanto asignamos esta constante de velocidad a un proceso de isomerización no relacionado con el proceso de desplegamiento y descartamos, al igual que en los otros dominios, la presencia de intermediarios en la desnaturalización de E2C-45.

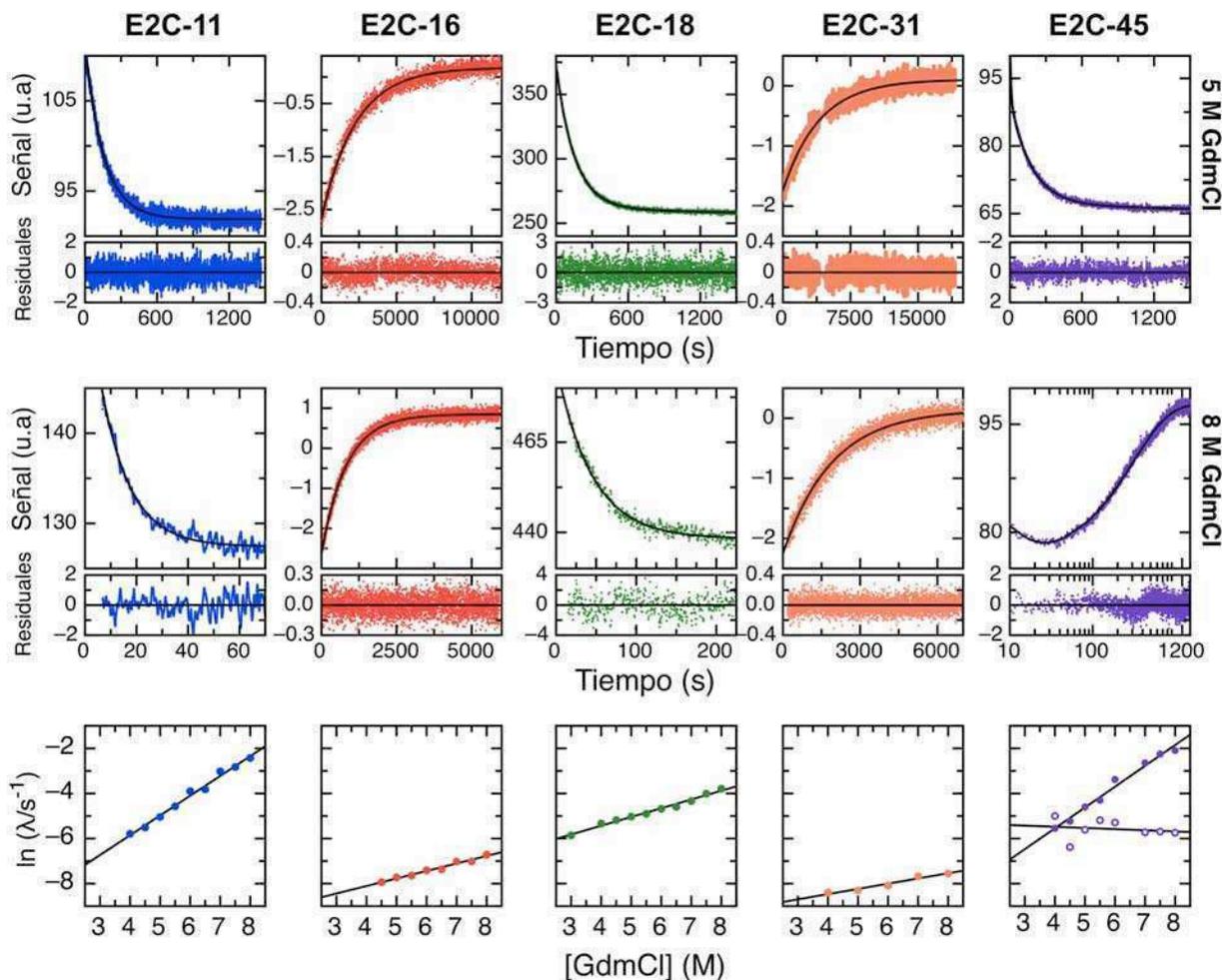


Figura IV.4. Cinética de desnaturalización de E2C-11, E2C-16, E2C-18, E2C-31 y E2C-45 en 50 mM fosfato de sodio pH 7, 1 mM DTT a 25°C, monitoreada por CD y fluorescencia. Primera y segunda fila: cinéticas de desnaturalización inducidas por 5 y 8 M de Gdm.Cl respectivamente. Las cinéticas para las homólogas E2C-16 y E2C-31 fueron monitoreadas por la señal de dicroísmo circular a 225 nm (mdeg) y las cinéticas para los dominios E2C-11, E2C-18 y E2C-45 fueron monitoreadas por fluorescencia de triptófanos (unidades arbitrarias). La concentración de proteica fue de 1 μ M en los experimentos de dicroísmo circular y 0.5 μ M en los experimentos de fluorescencia. La línea sólida corresponde al ajuste a una exponencial simple y doble en el caso de E2C-45, debajo de cada curva mostramos los residuales de ajuste. Tercera fila: dependencia de la constante de velocidad observada para la desnaturalización de E2C, gráfico de Chevron. La recta corresponde a un ajuste lineal.

En procesos de plegamiento de dos estados, el logaritmo de la velocidad aparente observada λ en función de la concentración de agente desnaturalizante conforma una curva en

forma de V o gráfico de Chevron (Ikai 1971; Matthews 1987; Kiefhaber *et al.* 2008). Los gráficos representados en la **Figura 4 tercera fila**, corresponden a gráficos parciales de Chevron a altas concentraciones de Gdm.Cl. Por ello se observa solamente el brazo de la ‘V’ correspondiente a la reacción de desplegamiento. En todos los casos, el logaritmo neperiano de la constante de velocidad observada depende linealmente con la concentración de Gdm.Cl, lo que es compatible con un proceso de dos estados y nos permite extrapolar el valor de la constante de desplegamiento en ausencia de desnaturalizante. En la **Tabla 3**, presentamos los parámetros del ajuste a un modelo de dos estados de las cinéticas de desnaturalización (ver métodos).

Tabla IV.3. Parámetros de las cinéticas de desnaturalización.

Proteína	$\ln(k_u/s^{-1})^a$	$1/k_u$ (horas)	m_u^b (kcal·mol ⁻¹ ·M ⁻¹)	α_D^c
E2C-11	-9.36 ± 0.22	3.2 ± 0.1	0.52 ± 0.02	0.11 ± 0.01
E2C-16	-9.44 ± 0.14	3.5 ± 0.1	0.20 ± 0.01	0.05 ± 0.01
E2C-18	-6.93 ± 0.13	0.3 ± 0.1	0.23 ± 0.01	0.04 ± 0.01
E2C-31	-9.64 ± 0.24	4.3 ± 0.2	0.16 ± 0.02	0.04 ± 0.01
E2C-45	-9.24 ± 0.30	2.9 ± 0.1	0.55 ± 0.03	0.14 ± 0.01

^a $\ln k_u$ corresponde a la ordenada al origen del ajuste de un modelo de dos estados a los datos. ^b constante de proporcionalidad, pendiente del ajuste de un modelo de dos estados a los datos. ^c calculado de acuerdo

$$\alpha_D = \frac{\partial \Delta G_u^0 / \partial [Gdm.Cl]}{\partial \Delta G_{eq}^0 / \partial [Gdm.Cl]} = \frac{m_u}{m_{eq}} \text{ (Tanford 1970), utilizando los datos medidos al equilibrio, Tabla 2.}$$

Comparando entre las distintas proteínas, podemos notar en los ejemplos de las cinéticas mostradas en la **Figura 4**, que a una misma concentración de Gdm.Cl los tiempos de finalización de la reacción varían notablemente entre los dominios. Adicionalmente, la pendiente de las rectas o m_u varía (**Tabla 3 y Figura 4, última fila**). Llamativamente salvo para E2C-18, el logaritmo neperiano de k_u en ausencia de Gdm.Cl es prácticamente el mismo para las E2C con un valor promedio de -9.4. Esto indica cualitativamente que la vida media de los dominios o estabilidad cinética en ausencia de agente desnaturalizante es alta, con un valor promedio de 4.3 horas para E2C-11, E2C-16, E2C-31 y E2C-45 (**Tabla 3**). E2C-18 difiere del grupo, siendo el dominio cinéticamente menos estable de los cinco. Este dato resultó sorprendente ya que es el dominio de mayor estabilidad al equilibrio (**Tabla 2**), sugiriendo que en este caso la estabilidad de E2C-18

se explica más por un rápido plegamiento que por la estabilidad cinética.

La determinación de la dependencia de la k_u con la concentración de Gdm.Cl, m_u , facilita la exploración de ciertas propiedades estructurales de la barrera de energía libre o estado de transición para la reacción de desnaturalización de los diferentes dominios E2C. Como mencionamos en el estudio al equilibrio el parámetro m_{eq} es experimentalmente proporcional al cambio en área accesible al solvente para la reacción (ΔASA) (Myers *et al.* 1995). Podemos postular que el parámetro m_u es también proporcional a ΔASA . Esto nos permite comparar el ΔASA entre los estados nativo y desplegado (proporcional a m_{eq}) con el ΔASA entre los estados nativo y de transición para la reacción de desnaturalización (proporcional a m_u). Esta comparación se realiza mediante el parámetro α_D , definido como $\alpha_D = m_u/m_{eq}$. Valores de α_D cercanos a 1 indican que el estado de transición de la reacción posee las mismas propiedades que el estado desplegado, mientras que si $\alpha_D = 0$ el estado de transición posee las mismas propiedades que el estado nativo. En las cinco proteínas homólogas el α_D es cercano a cero (0.08 ± 0.05 en promedio) indicando que la posición relativa del estado de transición en las coordenadas de la reacción se acerca al estado nativo en todos los dominios (**Tabla 3**). Podemos concluir que solo un ~8 % del cambio en la exposición de área al solvente (ΔASA) del estado nativo al desplegado ya ocurrió en el estado de transición, una mínima perturbación estructural en el estado de transición. En conjunto con los experimentos al equilibrio, podemos concluir que a pesar de la extensa divergencia de secuencia, las cinco variantes naturales de E2C son homodímeros altamente estables, cuya disociación es una simple reacción de dos estados, siendo el paso limitante cinético muy similar entre ellas y con un estado de transición similar al estado nativo en términos de su interacción con el Gdm.Cl. A continuación estudiamos las diferencias funcionales entre las cinco proteínas, focalizándonos en la capacidad de discriminación de secuencia nucleotídica.

Interacción entre E2C y BS2, cambios conformacionales en el ADN.

El estudio del reconocimiento de secuencia nucleotídica fue abordado para las cinco proteínas homólogas E2C mediante la técnica de ITC, absorbancia y CD en las mismas condiciones fisiológicas que fijamos en el Capítulo III, es decir 200 mM fosfato pH 7 a 25°C. Como mencionamos anteriormente, el complejo entre E2C-16 y el sitio de ADN BS2 es el más explorado biofísicamente en la literatura (Ferreiro *et al.* 2000; Ferreiro 2003; Dellarole 2005;

Ferreiro *et al.* 2005; Cicero *et al.* 2006; Dellarole *et al.* 2007; Ferreiro *et al.* 2008; Eliseo *et al.* 2009; Sanchez *et al.* 2010), y Capítulo III. Por lo tanto, considerando que el estudio de la interacción E2C-ADN es inédita para varios dominios de los que estamos estudiando, analizamos primero la formación del complejo E2C-ADN para el sitio BS2 a nivel conformacional (**Figura 5**).

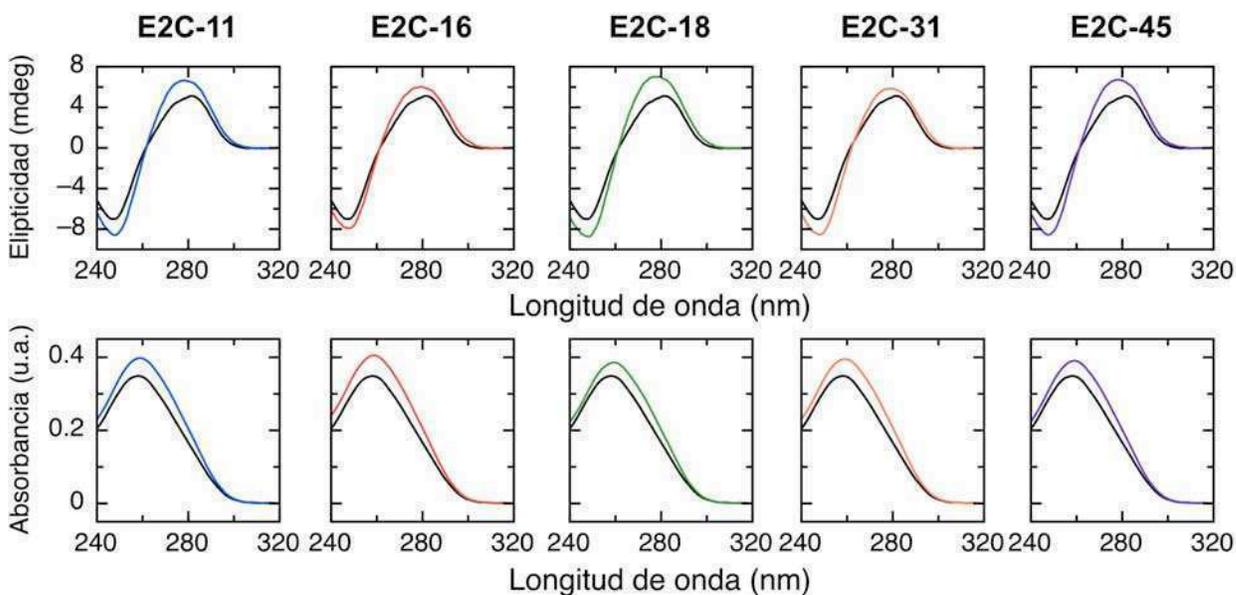


Figura IV.5. Cambios conformacionales en BS2 al interactuar con E2C-11, E2C-16, E2C-18, E2C-31 o E2C-45. Primera fila: espectros de CD de BS2 libre (negro) y unido (coloreado) para cada dominio. **Segunda fila:** espectros de absorbancia de BS2 libre (negro) y unido (coloreado) para cada dominio. A los espectros de los complejos se les restó el espectro de la proteína libre correspondiente. Todos los experimentos fueron realizados en el tampón químico 200 mM de fosfato de sodio pH 7, 0.2 mM de DTT a 25°C en una estequiometría E2C:ADN de 1:1 a 15 μ M. Ambas filas fueron llevadas a la misma escala para facilitar la comparación.

De acuerdo a evidencias cristalográficas, la formación de los complejos E2C-ADN induce una torsión considerable del ADN hacia el surco menor (ver por ejemplo la **Figura II.1**) y mínimamente acomodamientos en la proteína (Kim *et al.* 2000; Hooley *et al.* 2006). Los cambios conformacionales del ADN al unir E2C, fueron también reportados mediante métodos espectroscópicos (Lima y Prat Gay 1997; Ferreiro *et al.* 2000). Los métodos espectroscópicos de CD y absorbancia, reportan principalmente cambios conformacionales que tienen lugar en el ADN ya que el coeficiente de extinción del ADN supera ampliamente al de las proteínas (ver métodos), gobernando entonces la señal espectral (Gray *et al.* 1995; Gray 1996). En la **Figura 5**, mostramos los espectros del sitio de ADN libre BS2 (líneas negras) y de los complejos binarios E2C:ADN (líneas coloreadas de acuerdo al homólogo) a los cuales se les sustrajo los espectros de las proteínas libres.

En primer lugar, tanto los espectros de CD del ADN libre como unido presentan un pico positivo a 275 nm, un cruzamiento del cero a 260 nm y un pico negativo a 245 nm (**Figura 5, primera fila**), en todos los dominios E2C medidos. Estas tres propiedades espectrales son características de ADN en forma-B (Baase *et al.* 1979; Gray *et al.* 1995; Gray 1996). Concluimos que la unión de E2C a BS2 no modifica la naturaleza de forma-B del BS2 en ningún caso, en excelente acuerdo con datos cristalográficos del ADN de unión a E2 libre (Rozenberg *et al.* 1998; Hizver *et al.* 2001) y unido a E2C (Kim *et al.* 2000; Hooley *et al.* 2006). En segundo lugar, observamos un incremento en ambas señales espectroscópicas (CD y absorbancia), en la formación de los cinco complejos binarios. Dicha diferencia espectral es muy similar entre las distintas proteínas homólogas (**Figura 5**), sugiriendo que los cambios inducidos por la unión de E2 en la estructura del BS2 son semejantes. Este resultado se encuentra también de acuerdo con resultados anteriores observados para los dominios E2C-6, E2C-16 y E2C-18 (Lima y Prat Gay 1997; Ferreira *et al.* 2000; Kim *et al.* 2000; Hooley *et al.* 2006). A modo de abordar la capacidad de discriminación de secuencia, a continuación medimos la termodinámica de la unión de las cinco proteínas homólogas a tres sitios de ADN: BS1, BS2 y BSR.

Interacción entre E2C y ADN, termodinámica de unión.

Llevamos a cabo el análisis de la interacción de E2C-11, E2C-16, E2C-18, E2C-31 y E2C-45 y los sitios de ADN BS1, BS2 y BSR por ITC en 200 mM de fosfato de sodio pH 7 a 25°C (**Figura 6**). Los sitios de ADN BS1 y BS2 que estudiaremos pertenecen al genoma del papilomavirus de tipo 16 y fueron elegidos en base a su importancia biológica, como observamos en el Capítulo anterior y ahondaremos en el próximo Capítulo. Por otro lado, el sitio BSR, de sitio 'Recto', fue diseñado racionalmente mutando los nucleótidos del espaciador típicamente rico en adeninas y timidinas, como vimos para todos los sitios de ADN diana de E2 de alfapapilomavirus (ver **Figuras II.3 y 5 Introducción**), a citosinas y guaninas. De esta forma diseñamos un sitio de unión a E2 que pierde la pre-curvatura del surco menor (**Figura II.1**), responsable de la lectura indirecta (Shatzky-Schwartz *et al.* 1997; Crothers 1998; Hizver *et al.* 2001; MacDonald *et al.* 2001; Barbic *et al.* 2003; Xi *et al.* 2010), pero que mantiene las bases responsables de la lectura directa de la interacción E2C-ADN (Prat Gay *et al.* 2008). Las secuencias de los oligonucleótidos se encuentran detalladas en la **Tabla III.2**.

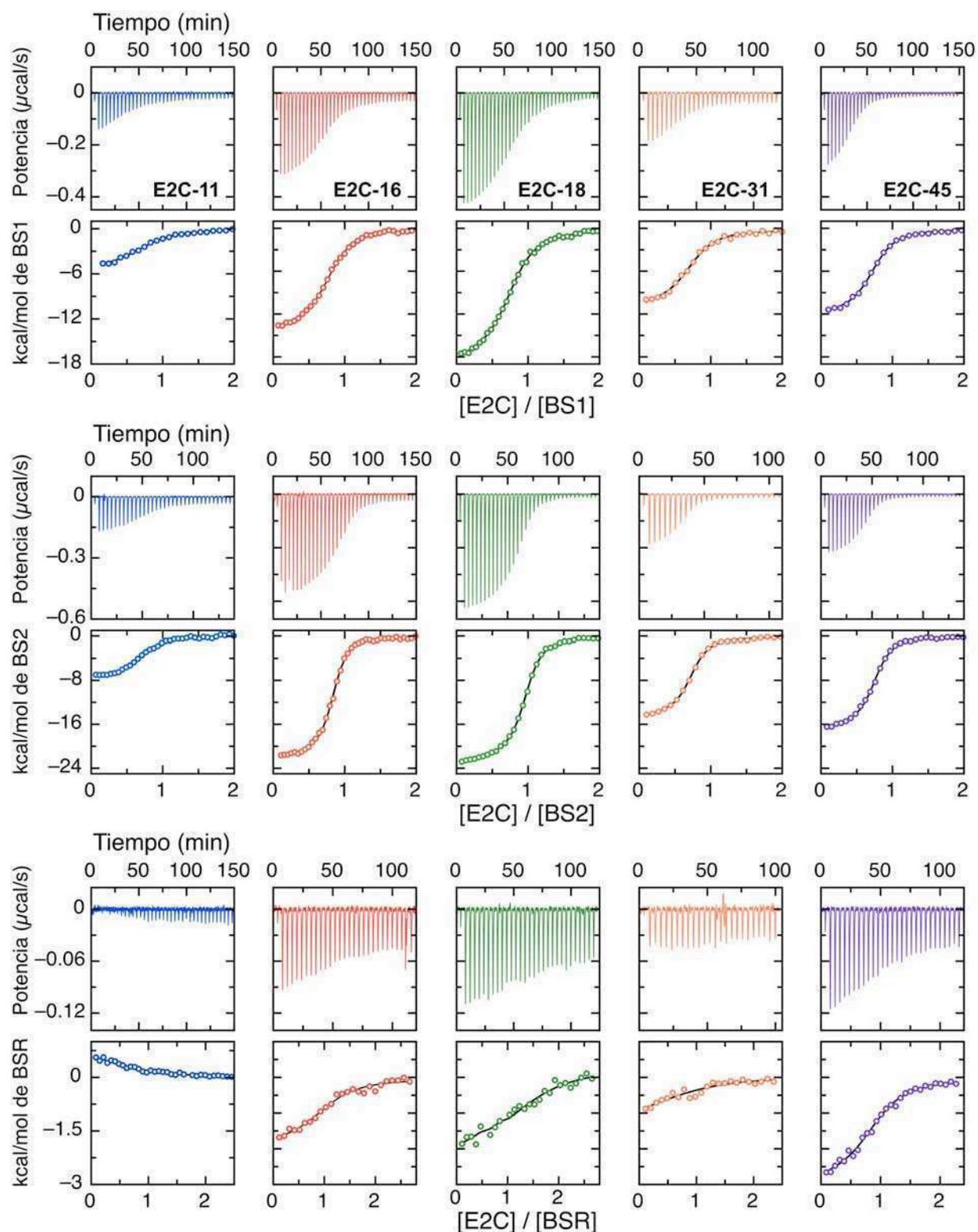


Figura IV.6. Unión de E2C a ADN. ITC para la unión de E2C-11, E2C-16, E2C-18, E2C-31o E2C-45 (columnas) a BS1 (Primera fila), BS2 (Segunda fila) y BSR (Tercera fila). Para cada sitio mostramos los datos crudos y los datos integrados y normalizados por concentración correspondientes. Las líneas negras corresponden al ajuste de un modelo de un sitio de unión a los datos de ITC. Tampón químico de medición: 200 mM de fosfato de sodio pH 7, 0.2 mM de DTT a 25°C. Cada fila fue llevada a la misma escala para facilitar la comparación entre dominios.

Tabla.IV.4. Termodinámica de la interacción entre los dominios E2C y ADN.

Sitio	Proteína	ΔH (kcal/mol)	K_D (nM)	ΔG (kcal/mol)	$-T \cdot \Delta S$ (kcal/mol)	N	C
BS1	E2C-11	-5.8±0.7	579±181	-8.51±0.18	-2.7±0.5	0.7±0.1	10±4
	E2C-16	-15.2±0.4	248±29	-9.01±0.07	6.2±0.3	0.8±0.1	26±5
	E2C-18	-19.5±0.9	399±17	-8.73±0.02	10.8±0.8	0.8±0.1	19±1
	E2C-31	-12.4±0.2	192±12	-9.16±0.04	1.72±0.2	0.7±0.1	17±1
	E2C-45	-12.3±0.2	254±61	-9.01±0.14	3.3±0.3	0.7±0.1	18±5
BS2	E2C-11	-7.9±0.2	216±32	-9.09±0.08	-1.2±0.1	0.7±0.1	21±2
	E2C-16	-22.7±0.4	90±40	-9.61±0.24	13.1±0.2	0.8±0.1	83±24
	E2C-18	-24.1±0.3	127±3	-9.41±0.1	14.7±0.3	0.9±0.1	60±4
	E2C-31	-14.8±0.1	92±4	-9.60±0.02	5.2±0.1	0.7±0.1	37±1
	E2C-45	-16.36±1.1	74±46	-9.73±0.39	6.6±1.5	0.7±0.1	50±24
BSR	E2C-11	0.8±0.2	2201±676	-7.72±0.14	-8.6±0.1	0.7±0.1	2±1
	E2C-16	-2.2±0.1	1070±233	-8.14±0.15	-6.0±0.3	0.9±0.2	9±2
	E2C-18	-3.1±0.4	1610±75	-7.90±0.04	-4.9±0.3	1.0±0.3	7±1
	E2C-31	-1.5±0.3	1973±1247	-7.78±0.09	-6.3±0.5	0.8±0.2	2±1
	E2C-45	-2.9±0.2	1045±397	-8.16±0.24	-5.8±0.2	0.9±0.1	8±4

El ajuste a un modelo de unión de un sitio de los datos crudos permite conocer el cambio en entalpía (ΔH) y la constante de disociación (K_D) de la unión. ΔG es calculado como $RT \cdot \ln(K_D)$, $-T \cdot \Delta S$ se calcula como ΔG menos ΔH . N es la estequiometría de la reacción y C es el relación entre la concentración de ligando y la constante de disociación (Wiseman *et al.* 1989). Los parámetros representados provienen de por lo menos dos mediciones ajustadas globalmente (Houtman *et al.* 2007), los errores corresponden al desvío estándar de los ajustes individuales.

En la **Figura 6** se ilustra una muestra de las quince titulaciones calorimétricas de la unión de los cinco dominios E2C (columnas) a los tres sitios de ADN (filas), mostrando en cada experimento los datos crudos y normalizados por concentración. El modelo de un sitio de unión fue ajustado globalmente a los datos de al menos dos titulaciones cuyos parámetros se detallan en la **Tabla 4**. La unión de las cinco E2C homólogas a los tres sitios de ADN presentan una estequiometría de reacción de 0.7 o mayor y un valor C por encima de 10, (**Tabla 4**), validando el ajuste del modelo de unión 1:1 a los datos. Es importante mencionar que el estudio de las condiciones de medición llevado a cabo con el sitio BS2 y la proteína E2C-16 (**Figura III.1**), resultó clave para poder medir correctamente por ITC la formación de quince complejos

diferentes. El ajuste nos provee la energías libres de unión, ΔG , y sus componentes entrópicos y entálpicos (**Tabla 4, Figura 7**).

Al igual que lo visto para E2C-16 en el Capítulo anterior, la unión a los sitios BS1 y BS2 por las cinco proteínas homólogas resultó entálpicamente dirigida para todos los dominios, siendo la unión al BS2 significativamente más afín que la unión al BS1, encontrándose diferencias de hasta tres veces en K_D (**Tabla 4**). Esta característica se encuentra en acuerdo con nuestra premisa previa que postula que la especificidad de secuencia se mantiene para todos los dominios de E2C de alfa papilomavirus, Capítulo II y (Sanchez *et al.* 2008). Sorprendentemente, la unión al sitio diseñado BSR es para los cinco dominios la de menor afinidad. Es decir, la jerarquía de afinidades a los tres sitios de ADN es la misma en los cinco dominios ensayados. El promedio de energía libre de unión, ΔG , de las cinco proteínas es: $\Delta G_{BS2} = -9.49 \pm 0.25$ kcal/mol $>$ $\Delta G_{BS1} = -8.88 \pm 0.26$ kcal/mol $>$ $\Delta G_{BSR} = -7.86 \pm 0.34$ kcal/mol (**Tabla 4 y Figura 7A**), evidenciando que la jerarquía de afinidades es significativa con un mínimo desvío estándar del ΔG promedio. Es notable como disminuye la afinidad para el sitio BSR, mostrando la relevancia funcional de la lectura indirecta en el sistema E2C-ADN.

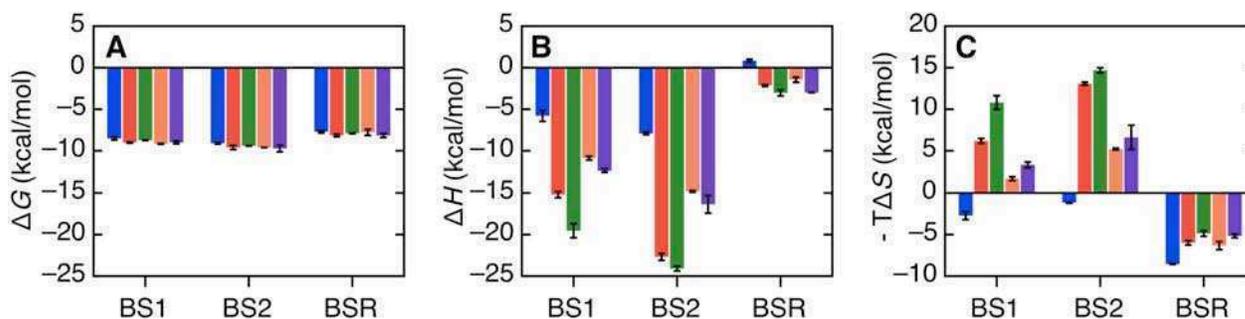


Figura IV.7. Parámetros para la unión de E2C a BS1, BS2 y BSR. (A) ΔG . (B) ΔH . (C) $-T\Delta S$. El código de color de las barras corresponde a un dominio distinto: azul E2C-11, rojo E2C-16, verde E2C-18, naranja E2C-31 y violeta E2C-45.

Sin embargo, la diferencia de energía libre de Gibbs para cada sitio entre las distintas proteínas homólogas no varía en la misma escala que la diferencia de entalpía y entropía. En la **Figura 6** podemos apreciar visualmente que el parámetro termodinámico de diferencia de entalpía (ΔH), correspondiente a la altura de la transición en los datos normalizados por concentración y expresado en kcal/mol, varía apreciablemente tanto entre sitios y para la misma proteína (**Figura 6, columnas**) como entre proteínas y para un mismo sitio (**Figura 6, filas**). En efecto, el promedio de ΔH para las cinco proteínas es $\Delta H_{BS2} = -17.17 \pm 6.52$ kcal/mol $>$ $\Delta H_{BS1} = -$

12.75±5.11 kcal/mol > $\Delta H_{\text{BSR}} = -1.72 \pm 1.67$ kcal/mol evidenciando el gran desvío estándar que hay para la unión a los tres sitios (**Figura 7B** y **Tabla 4**). Es llamativo como en particular la entalpía de la reacción disminuye abruptamente para la unión al BSR, siendo hasta entálpicamente desfavorable para el dominio E2C-11 (**Figura 7B** y **Tabla 4**). Podemos apreciar este mismo comportamiento en el caso de la diferencia de entropía ($-T\Delta S$), dónde son destacables dos particularidades, primero: la unión de E2C-11 a los tres sitios de ADN es siempre entrópicamente favorable; y segundo: la unión de todas las proteínas al sitio BSR es gobernada por la entropía (**Figura 7C** y **Tabla 4**). Esta peculiaridad puede deberse a la falta de pre-curvatura del ADN en el estado libre, a la ausencia de lectura indirecta en el complejo E2C-BSR, por lo tanto nos puede dar indicios de qué evento gobierna la magnitud del ΔH para la interacción E2C-ADN (ver Discusión).

Tabla.IV.5. Diferencias energéticas y discriminación de secuencia.

Δ Sitio	Proteína	$\Delta\Delta G$ (kcal/mol)	$\Delta\Delta H$ (kcal/mol)	$-T \cdot \Delta\Delta S$ (kcal/mol)	Especificidad ^a
BS(2-I)	E2C-11	-0.58±0.20	-2.13±0.69	-1.55±0.49	2.7±0.1
	E2C-16	-0.60±0.25	-7.49±0.56	-6.89±0.35	2.8±0.1
	E2C-18	-0.68±0.03	-4.54±0.89	-3.87±0.88	3.1±0.1
	E2C-31	-0.43±0.03	-3.93±0.25	-3.49±0.25	2.1±0.1
	E2C-45	-0.73±0.42	-4.03±1.08	-3.29±1.49	3.5±0.1
BS(2-R)	E2C-11	-1.38±0.17	-8.76±0.23	7.39±0.08	10±0.1
	E2C-16	-1.47±0.28	-20.54±0.43	19.08±0.31	12±0.1
	E2C-18	-1.51±0.04	-21.06±0.47	19.55±0.46	13±0.1
	E2C-31	-1.81±0.38	-13.35±0.33	11.54±0.5	21±0.1
	E2C-45	-1.57±0.46	-13.35±1.07	-11.79±1.47	14±0.2
BS(1-R)	E2C-11	-0.79±0.23	-6.63±0.7	5.84±0.49	3.8±0.1
	E2C-16	-0.87±0.17	-13.05±0.39	12.18±0.40	4.3±0.1
	E2C-18	-0.83±0.04	-16.51±0.93	15.69±0.89	4.1±0.1
	E2C-31	-1.38±0.38	-9.43±0.39	8.05±0.54	10.3±0.1
	E2C-45	-0.84±0.28	-9.34±0.18	8.5±0.39	4.1±0.1

^a Especificidad de secuencia calculada de acuerdo a la relación entre las K_D involucradas en cada par, por ejemplo $K_{\text{DBS2}}/K_{\text{DBS1}}$. Los errores fueron propagados en todos los casos a partir de los datos de la **Tabla 5**.

La discriminación termodinámica de secuencia puede ser analizada con mayor claridad calculando la diferencia de los parámetros termodinámicos ($\Delta\Delta G$, $\Delta\Delta H$ y $-\Delta T\Delta S$) entre dos sitios distintos para un dominio determinado (**Tabla 5** y **Figura 8**). Es decir, la diferencia para los pares de sitios BS2 y BS1 (BS1-2), BS2 y BSR (BS1-R) y BS1 y BSR (BS1-R). El gráfico de $\Delta\Delta G$, (**Figura 8A**) muestra claramente la estricta conservación de reconocimiento de ADN por las cinco proteínas, con $\Delta\Delta G$ casi idénticos para cada sitio de ADN (**Tabla 5**). En cuanto a los componentes entálpicos y entrópicos, la variación energética es mayor y no sigue un patrón determinado (**Figura 8B** y **8C**). En el reconocimiento de ADN por E2C, el índice de especificidad de secuencia, una forma de cuantificar la discriminación de secuencia utilizando el cociente de K_D , se encuentra íntimamente relacionado con la plasticidad o flexibilidad proteica (Dellarole *et al.* 2007; Brown *et al.* 2010). Los mismos pares analizados arriba fueron estudiados en cuanto a la especificidad de secuencia (**Tabla 5**). Para cada par de sitios, en promedio, la especificidad es similar entre los dominios: 3 veces más afín al sitio BS2 que al BS1, 10 veces más afín al sitio BS2 que al BSR y 4 veces más afín al sitio BS1 que al BSR (**Tabla 5**). La única diferencia reside en la especificidad de unión al BSR del dominio E2C-31, sugiriendo diferencias frente a las otras cuatro proteínas E2C en cuanto a plasticidad para reconocer un sitio con la lectura indirecta anulada.

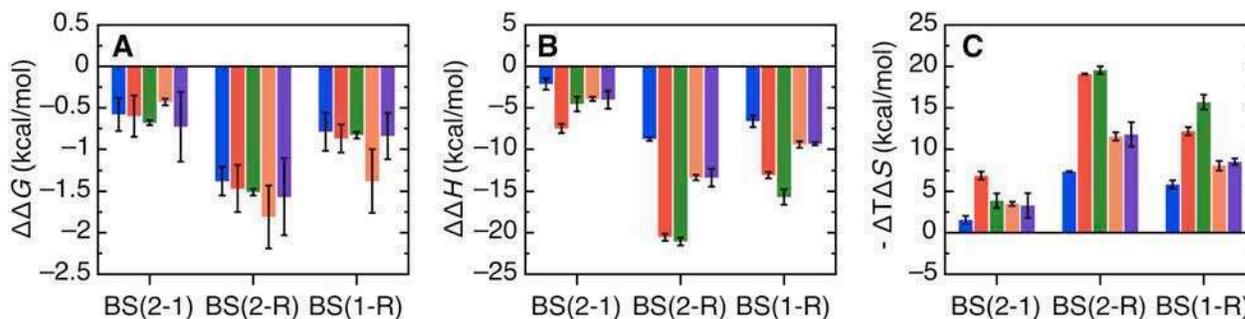


Figura IV.8. Análisis diferencial de los parámetros para la unión de E2C a BS1, BS2 y BSR. Diferencias energéticas entre la unión de E2C a BS1, BS2 y BSR. (A) $\Delta\Delta G$ (B) $\Delta\Delta H$ (C) $-\Delta T\Delta S$ correspondientes con la unión de E2C a BS1 y BS2, a BS2 y BSR y a BS1 y BSR. El código del color de las barras corresponde a un dominio distinto: azul E2C-11, rojo E2C-16, verde E2C-18, naranja E2C-31 y violeta E2C-45.

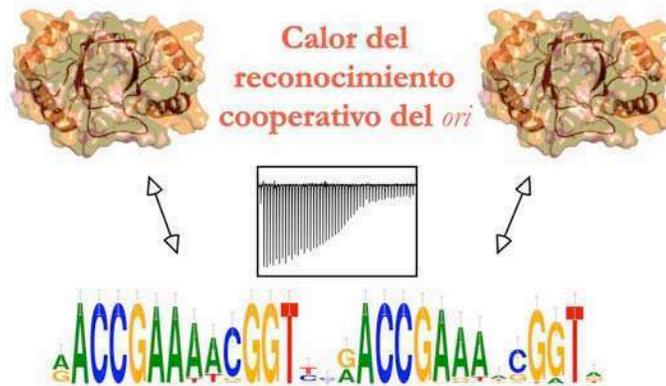
En forma global describimos que existen considerables diferencias en la disección de la energía libre de unión en cuanto a las contribuciones entálpicas y entrópicas entre las cinco proteínas. Así como intentamos correlacionar el m_{eq} con diferentes parámetros, buscamos correlacionar curzando datos funcionales con datos de estabilidad pero no encontramos ninguna

asociación que nos permita racionalizar, por ejemplo, la amplia variación entálpica entre los cinco dominios homólogos. Los componentes tanto entálpicos como entrópicos se originan a partir del balance de múltiples eventos propios de la formación del complejo proteína-ADN. En cuanto a la entalpía de la reacción, el calor de la desolvatación en las superficies polares; las interacciones directas: uniones de hidrógenos entre bases y residuos y mediados por aguas o iones, pares iónicos y contactos no polares; las interacciones indirectas: cambios conformacionales locales y/o globales, torsiones, giros, aperturas del ADN, eventos de protonación, entre otros. En cuanto a la entropía de la reacción, la desolvatación de superficies no-polares; la redistribución de contraiones; los cambios en grados de libertad y rotación de las moléculas involucradas; impedimentos estéricos y restricciones de salida de moléculas de agua ‘atrapadas’ en la interfaz del complejo, entre otros. Por lo tanto, en el equilibrio, todos estos sucesos aportan a los valores termodinámicos medidos. Es por ello que la interpretación en forma aislada de las diferencias en la interacción entre los tres sitios y las cinco proteínas homólogas se vuelve muy compleja. Sin embargo, el haber utilizado un oligonucleótido diseñado especialmente para anular el componente de lectura indirecta puede dar indicios sobre la naturaleza de las energías medidas y el haber medido la interacción E2C-ADN en treinta condiciones distintas variando el entorno químico (Capítulo III, (Eliseo *et al.* 2009)), la secuencia del ADN (Capítulo III y IV, (Dellarole *et al.* 2010)), la secuencia de la proteína (presente Capítulo y (Cicero *et al.* 2006)) y la naturaleza dimérica (Dellarole *et al.* 2007), puede iluminar aspectos generales de la termodinámica de la interacción E2C-ADN. Dado que todos estos datos fueron obtenidos más allá del presente Capítulo, abordaremos un análisis integrador en la Discusión.

La conservación de la estabilidad y reconocimiento de ADN a costas de una extensa divergencia de secuencia en cinco dominios homólogos nos motivó a investigar sistemas más complejos con el fin de estudiar nuevas ventajas evolutivas residentes en la extensa variabilidad de secuencia de E2C del *genus* alfa. Los datos funcionales expuestos en el presente Capítulo nos permitirán estudiar la termodinámica de la interacción de un sistema con un grado mayor de complejidad. Estudiaremos la unión de los sitios BS1 y BS2 dispuestos en tándem (DBS, doble sitio de unión), para tres de los cinco dominios homólogos, buscando comprender las bases termodinámicas de la unión cooperativa en este bloque funcional clave en el desarrollo del ciclo infeccioso del virus.

Capítulo V

“Bases termodinámicas del reconocimiento cooperativo por E2 del origen de replicación y transcripción de papilomavirus”



*Le simple est toujours faux.
Ce qui ne l'est pas est inutilisable.*
Paul Valéry, Œuvres II

La unión cooperativa de múltiples proteínas a sitios de ADN adyacentes permite regulaciones precisas y económicas de cambios genéticos, del reclutamiento de la maquinaria transcripcional y de replicación (Ptashne 2004). En este Capítulo final estudiaremos la termodinámica de la interacción entre tres dominios E2C y los sitios BS1 y BS2 dispuestos en tándem (DBS), un sistema clave en la biología de alfafapilomavirus.

Abordaje experimental y modelo de unión.

El reconocimiento del DBS por E2 de HPV16 ocurre con leve cooperatividad positiva, de acuerdo a estudios por ensayos de retardo en gel en la movilidad electroforética (EMSA) (Tan *et al.* 1994). Sin embargo, para el modelado de los datos medidos por EMSA, las afinidades de E2 a los sitios BS1 y BS2 aislados se asumieron idénticas (Tan *et al.* 1994). En la presente tesis establecimos que la diferencia de afinidad entre BS1 y BS2 es superior a 0.6 kcal/mol (**Tabla IV.5**), en acuerdo con resultados medidos en otros laboratorios (Thain *et al.* 1997). Sin embargo consideramos que la técnica EMSA produce resultados cualitativos (Musheev *et al.* 2010). Estos antecedentes nos motivaron a estudiar la formación del complejo ternario mediante técnicas espectroscópicas e ITC. A continuación describiremos las ventajas que proveen tanto el ITC para la disección termodinámica de la cooperatividad en sus contribuciones entálpicas y entrópicas (Bains *et al.* 1991), como las técnicas espectroscópicas para adquirir información sobre los cambios estructurales asociados con la cooperatividad (Gray *et al.* 1995). Además, comparamos la unión al sitio DBS correspondiente al genoma de HPV16, entre los dominios E2C-16, 11 y 18 de forma tal de estudiar las consecuencias de la divergencia de secuencia para la unión cooperativa.

Como estudiamos en el Capítulo II, la secuencia nucleotídica y la distancia relativa de los sitios de unión de E2, BS1 y BS2 se encuentran altamente conservados en alfafapilomavirus. Adicionalmente, la secuencia nucleotídica y la distancia de los sitios de unión de Sp1 y TBP al DBS también se encuentran conservadas (**Figura 1A y 1B**), destacando la relevancia biológica de esta región reguladora en el ciclo de vida de alfafapilomavirus. Con el fin de visualizar la arquitectura molecular del sistema en estudio, construimos un esquema informativo del complejo ternario entre DBS y dos moléculas de E2C (**Figura 1C**). La representación molecular del sistema ternario posee una distancia entre los sitios BS1 y BS2 de una sola base, como ocurre en el genoma de HPV16, la secuencia elegida para estudiar termodinámicamente la interacción

ternaria (**Figura 1D**). Aún siendo cruda, esta aproximación estructural muestra claramente que la curvatura y torsión de los dos complejos binarios posiciona las dos moléculas de E2C en lados opuestos del ADN. Por lo tanto de acuerdo con este modelo, la unión aditiva de dos moléculas de E2C a DBS es en principio posible en el genoma de HPV16.

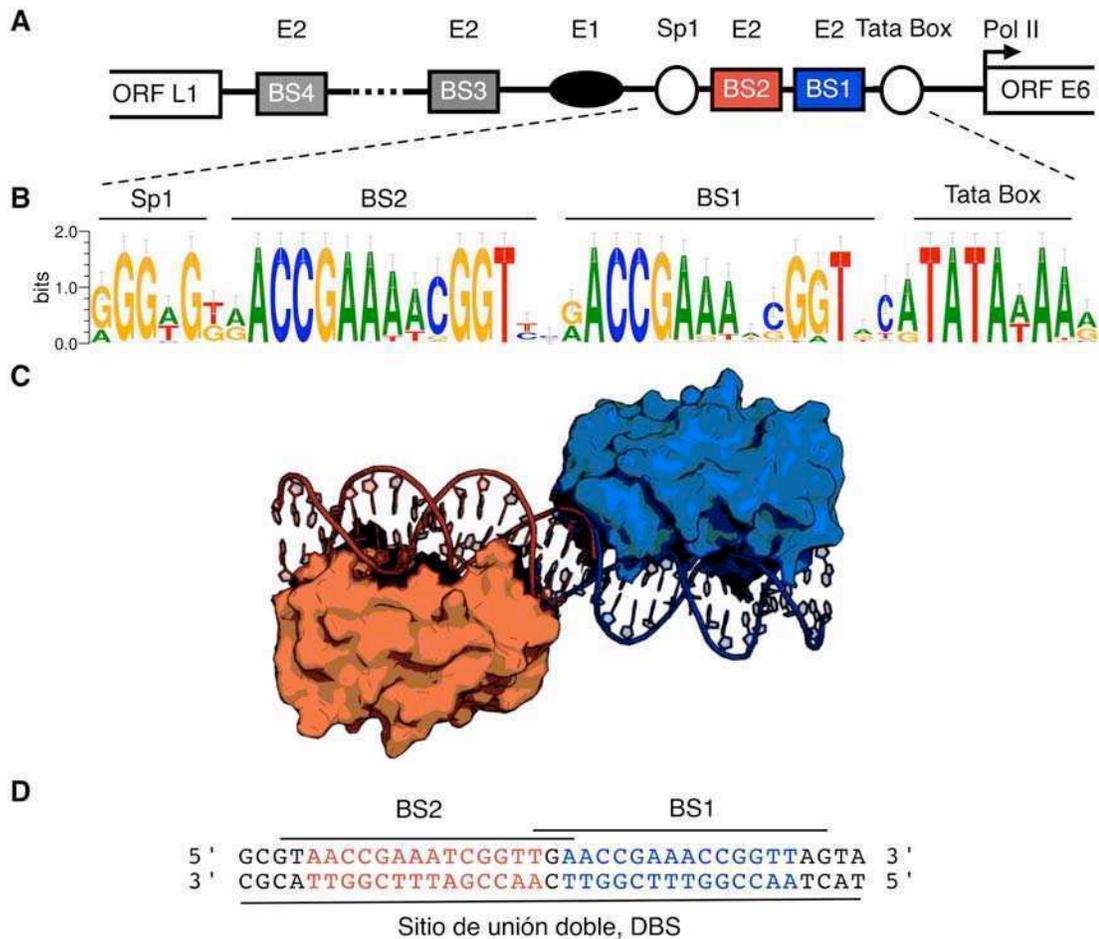
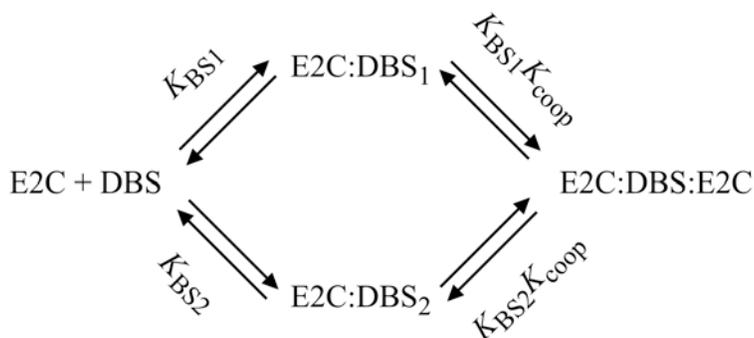


Figura V.1. Sitios de unión de E2 en tándem en la región reguladora de alpha papilomavirus. (A) Vista esquemática de la región reguladora de alpha papilomavirus. Se representan los marcos de lectura abiertos (ORF, por sus siglas en inglés) de L2 y E6, los sitios de unión de E2 BS1, BS2, BS3 y BS4, el sitio de unión de E2, el sitio de unión de Sp1 y la TATA box. (B) Logo de secuencias (Schneider y Stephens 1990; Crooks *et al.* 2004) de todos los elementos de respuesta *cis* del promotor de E6 en el *genus* alpha papilomavirus. (C) Modelo del complejo de E2C con dos sitios adyacentes separados por una base. El modelo fue construido alineando dos copias del complejo E2C:ADN de HPV-06 (2ayg) utilizando PyMol (Delano Scientific LLG). (D) Secuencia doble cadena del oligonucleótido DBS de HPV-16, con el BS1 coloreado en azul y el BS2 coloreado en rojo.

El primer paso para investigar el reconocimiento cooperativo de E2C al origen de replicación de HPV por ITC consistió en determinar las bases termodinámicas para el reconocimiento por E2C a cada uno de los sitios BS1 y BS2 por separado, que fue abordado en los Capítulos III y IV. La unión de E2C a los sitios BS1 y BS2 configurados en tándem y formando el sitio doble llamado DBS (**Figura 1D**), puede ser descrito de acuerdo al siguiente

mecanismo de unión general (Tan *et al.* 1994):



[1]

donde por un lado, E2C puede unir al BS1 dentro del DBS para formar el complejo binario E2C:DBS₁ con una constante de equilibrio K_{BS1} ; y por otro lado, E2C puede unir al sitio BS2 dentro del DBS para formar el complejo binario E2C:DBS₂ con una constante de equilibrio K_{BS2} . Finalmente, los complejos binarios se encuentran en equilibrio con el complejo ternario E2C:DBS:E2C cuya constante de equilibrio es $K_{BS2}K_{BS1}K_{coop}$. La constante adicional K_{coop} , por cooperatividad, puede ser equivalente a 1 si la formación del complejo ternario se da en forma aditiva o independiente. $K_{coop} > 1$ implica cooperatividad positiva, es decir la unión de la segunda molécula de E2C al DBS se ve favorecida cuando una molécula de E2C ya se encuentra unida al DBS. $K_{coop} < 1$ implica cooperatividad negativa, es decir, la unión de la segunda molécula de E2C al DBS se ve desfavorecida cuando una molécula de E2C ya se encuentra unida al DBS (Wyman y Gill 1990).

Inicialmente estudiamos la unión de E2C a DBS mediante las técnicas biofísicas de medición *in-vitro* de uso frecuente para investigar interacciones cooperativas (**Figura 2**). Es decir ensayos espectroscópicos de fluorescencia medidos por intensidad (**Figura 2A**) o anisotropía (**Figura 2B**, ver por ejemplo (Weinberg *et al.* 2004; Dragan *et al.* 2008)) y de retardo en la movilidad electroforética (EMSA) (**Figura 2A** y ver por ejemplo (Tan *et al.* 1994)). La interpretación de la interacción entre una proteína y dos sitios dispuestos en tándem se dificulta al aumentar el grado de complejidad del modelo de unión. Y el presente caso, donde los sitios aislados BS1 y BS2 no son ni equivalentes ni homogéneos, es decir, de distinta composición nucleotídica (**Figura 1D**) y afinidad individual (**Figura IV.6** y **Tabla IV.4**), presenta un elevado grado de complejidad. Por ello, no pudimos determinar la población de las especies formadas en las titulaciones seguidas mediante las técnicas mencionadas, aún luego de una extensa exploración de condiciones experimentales (datos no mostrados). Estudios previos del mismo sistema de interacción E2C-DBS medidos por EMSA simplificaron el modelo asumiendo

homogeneidad en los sitios aislados BS1 y BS2 (Tan *et al.* 1994), obteniendo de esta forma un valor de K_{coop} semi-cuantitativo con errores adicionales al error experimental (ver Discusión). Por lo tanto, ya que la discriminación de las poblaciones de todas las especies en nuestro sistema no fue posible por fluorescencia y EMSA, nos concentramos en ITC, donde observamos diferencias significativas en la señal de calor para complejo binario individual, E2C:BS1 y E2C:BS2 (Capítulos III y IV).

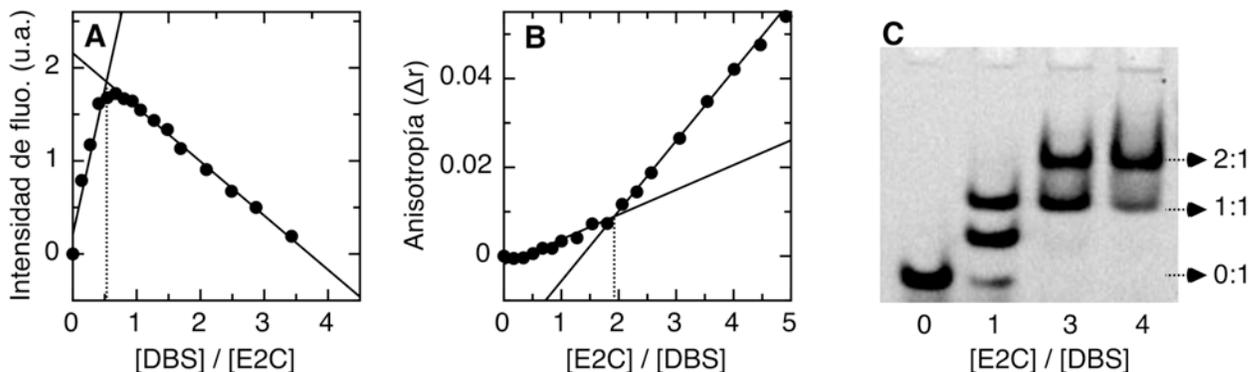


Figura V.2. Estequiometrías para la unión ‘reversa’ y ‘directa’ de E2C-16 a DBS seguida por fluorescencia y EMSA. (A) Ejemplo de titulación reversa, es decir titulación del ADN DBS a la proteína E2C a una concentración fija de 1 μ M, seguida por la intensidad de fluorescencia de triptofanos. (B) Ejemplo de titulación directa, es decir titulación de la proteína E2C al oligonuclótido DBS fluoresceinado a una concentración de 0.4 μ M, seguida por la anisotropía de fluorescencia de fluoresceína. (C) Ensayo de retardo en la movilidad electroforética titulando E2C al oligo DBS fluoresceinado en una concentración fija de 1 μ M y revelado en Storm 840. Tampón químico de medición: 200 mM NaCl, 10 mM de Bistris pH 7, 1 mM DTT a 25°C.

Calorimetría de titulación isotérmica para complejos ternarios.

Dada la naturaleza compleja de nuestro modelo de unión, no logramos cuantificar la unión cooperativa de E2C mediante técnicas de fluorescencia o EMSA. Sin embargo, con estas técnicas sí logramos determinar la estequiometría de la reacción (Figura 2). Como es de esperar para un sistema de dos sitios de unión (complejos ternarios), la estequiometría de la reacción varía según la naturaleza del titulante, a diferencia de sistemas de un sitio de unión (complejos binarios) que no varía la estequiometría, como analizamos anteriormente en la Figura III.2. Si el titulante es la molécula de ADN DBS, con dos sitios de unión para E2C (titulación directa), la estequiometría de la reacción es 0.5, correspondiente a la relación DBS:E2C 1:2 (Figura 2A); mientras que si el titulante es la molécula proteica E2C, con un solo sitio de unión para DBS (titulación reversa), la estequiometría de la reacción es 2, correspondiente a la relación E2C:DBS 2:1 (Figura 2B y 2C).

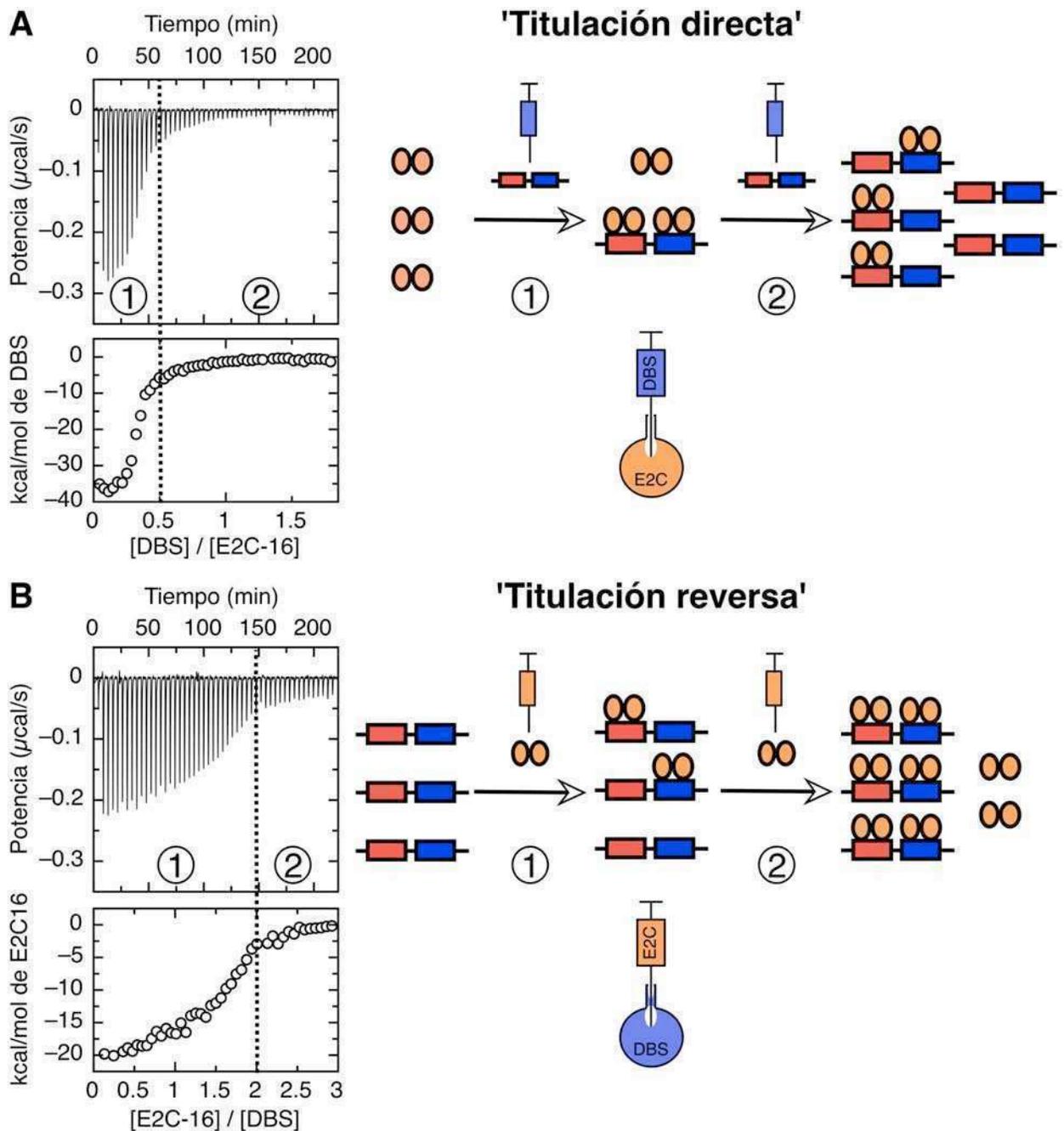


Figura V.3. Esquema descriptivo de titulaciones ‘directa’ y ‘reversa’ seguida por ITC para complejos ternarios. (A) Experimento de ‘titulación directa’, inyectando el sitio de ADN en tándem DBS a una celda conteniendo la proteína E2C. En las primeras inyecciones (①), el exceso de E2C conduce a la formación predominante de complejo ternario. En las últimas inyecciones (②), el exceso de DBS conduce a la formación predominante de complejos binarios. La estequiometría aparente de E2C:DBS:E2C es del orden de 0.5 (línea punteada). (B) Experimento de ‘titulación reversa’, inyectando la proteína E2C a una celda conteniendo el ADN DBS. En las primeras inyecciones (①), el exceso de ADN conduce a la formación predominante de complejos binarios. En las últimas inyecciones (②), el exceso de proteína conduce a la formación predominante de complejos ternarios. La estequiometría aparente de E2C:DBS:E2C es del orden de 2 (línea punteada).

En la **Figura 3** describimos gráficamente el desarrollo de titulaciones directas y reversas a modo de ilustrar la variación estequiométrica en experimentos de ITC. El panel **A** de la **Figura 3** muestra el experimento llevado a cabo por titulación directa, es decir inyectando ADN a una celda conteniendo la proteína. En las primeras inyecciones directas (①), el gran exceso de proteína conduce a la formación predominante del complejo ternario E2C:DBS:E2C. La estequiometría aparente DBS:proteína, $N_{directa}$, es del orden de $\frac{1}{2}$ y la entalpía aparente de unión (**Figura 3A**, aproximadamente de -38 kcal/mol) se encuentra cerca de la suma de los valores de ΔH para la unión de los sitios aislados BS1 y BS2 (**Tabla IV.2**). El panel **B** de la **Figura 3** muestra el experimento llevado a cabo por titulaciones reversas, es decir inyectando proteína a una celda conteniendo ADN. En las primeras inyecciones reversas (①), el gran exceso de ADN ahora conduce a la formación de los complejos binarios E2C:DBS₁ y E2C:DBS₂. La estequiometría aparente de proteína:ADN es ahora del orden de 2:1 y la entalpía aparente de unión es cerca del promedio de valores de ΔH para la unión de los sitios aislados BS1 y BS2 (comparar ΔH de -20 kcal/mol de la **Figura 3B** con los datos de la **Tabla III.2**). Por lo tanto, cualitativamente, las titulaciones calorimétricas de la interacción entre E2C y DBS presentaron la estequiometría esperada. Además, el hecho de que la magnitud de la señal para la formación del complejo ternario pueda ser interpretada de acuerdo a los valores medidos para los complejos binarios, es indicio de que la técnica facilitaría la discriminación de las poblaciones formadas. A continuación ajustaremos al modelo de dos sitios (esquema [1]), los datos medidos para cada proteína homóloga.

Termodinámica del reconocimiento cooperativo.

Con el fin de estudiar la interacción entre E2C y DBS por ITC, utilizamos las mismas condiciones de medición establecidas en la **Figura III.1** y utilizadas en la caracterización de los complejos binarios (**Figura IV.6**), es decir, 200 mM de fosfato de sodio pH 7, 0.2 mM de DTT a 25°C. De esta forma, los parámetros medidos para los complejos binarios serán útiles para interpretar cualitativa y cuantitativamente los datos de ITC para el complejo ternario entre E2C:DBS:E2C. Es importante recordar que la presencia de ADN no específico flanqueante a BS2 no afecta la constante de unión de E2C por BS2 (Sanchez *et al.* 2010), descartando posibles efectos coulómicos adicionales (Ballin *et al.* 2004). Además, las condiciones de fuerza iónica de medición (0.46 M dado por el fosfato de sodio) se encuentran muy por encima de las condiciones

favorables a eventos de unión adicionales inespecíficos (< 0.15 M, ver **Figura III.1** y **Tabla III.1**). Por estas razones, en las condiciones de medición, la constante K_{coop} no se vería influenciada por efectos electrostáticos secundarios, validando la utilización de los parámetros de unión de E2C a los sitios aislados BS1 y BS2 para interpretación del sistema DBS. La caracterización termodinámica del comportamiento de unión al sitio natural DBS de la proteína E2C de HPV-16 fue además realizada para las proteínas homólogas de HPV de tipo 11 y 18. De esta forma, además de control interno de validación del modelo, podremos estudiar la variabilidad de secuencia para el reconocimiento de dos sitios de ADN consecutivos, un nivel de complejidad mayor al analizado en el Capítulo IV.

Hemos ajustado en forma individual y global (**Tabla 1**) al modelo de unión de dos sitios, esquema [1], los datos de las titulaciones directas y reversas de la interacción entre E2C-16 (**Figura 4, primera columna**) E2C-11 (**Figura 4, segunda columna**) o E2C-18 (**Figura 4, tercera columna**) y el sitio de ADN DBS (**Figura 1D**). El ajuste de los datos medidos para cada proteína homóloga se llevó a cabo manteniendo fijos los parámetros para la unión a los sitios aislados que componen el complejo ternario, es decir las constantes de unión K_{BS1} y K_{BS2} y las entalpías ΔH_{BS1} y ΔH_{BS2} (ver **Tabla IV.4**). El ajuste produce los parámetros para la constante de afinidad adicional para la unión del complejo ternario llamada constante de cooperatividad K_{coop} (y por lo tanto, la energía libre de la cooperatividad ΔG_{coop}). El ajuste, por tratarse de mediciones calorimétricas, produce además los componentes entálpicos (ΔH_{coop}) y entrópicos (ΔS_{coop}) adicionales para formación del complejo ternario. La estequiometría de la reacción se deriva del parámetro de fracción incompetente de la reacción (Houtman *et al.* 2007), que corresponde a la concentración del titulante que no forma parte de la reacción; dicha fracción nos reporta entonces la desviación de la estequiometría del modelo de unión (Houtman *et al.* 2007). En la **Tabla 1** se encuentran los parámetros obtenidos de ajustes individuales y globales, es decir el ajuste al modelo de dos sitios de unión de los datos de una titulación ‘directa’ con una ‘reversa’ (ver **Figura 4** línea de ajuste y **Tabla 1** parámetros en gris).

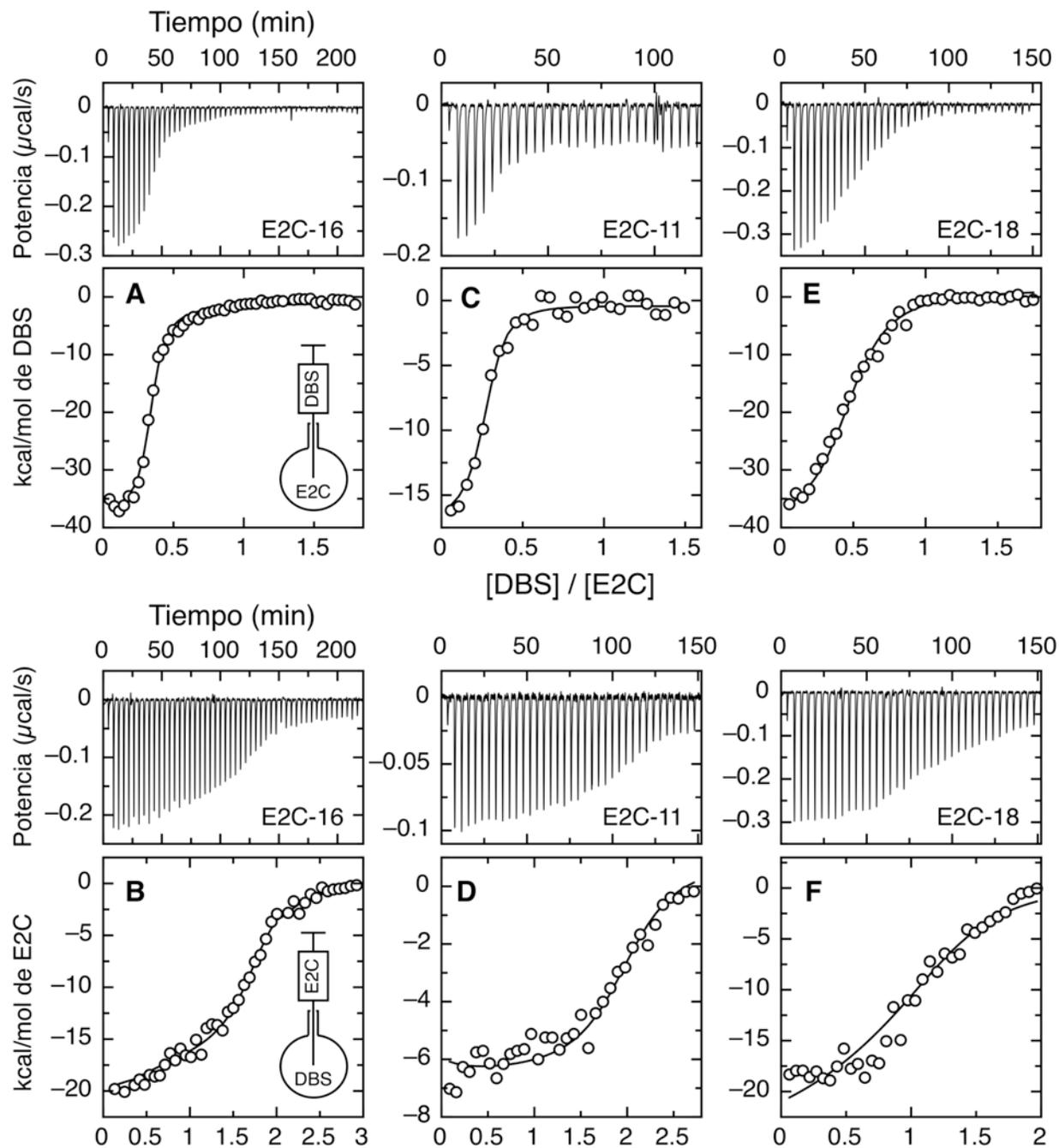


Figura V.4. Unión de E2C al ADN en tándem DBS. Cada panel muestra los datos de ITC crudos e integrados y normalizados por concentración. **Paneles de arriba:** titulaciones de ITC ‘directa’, es decir inyectando DBS en una celda conteniendo la proteína E2C para (A) E2C-16, (C) E2C-11 y (E) E2C-18. **Paneles de abajo:** titulaciones de ITC ‘reversas’, es decir inyectando E2C a una celda conteniendo DBS, siendo (B) E2C-16, (D) E2C-11 y (F) E2C-18. Las líneas se corresponden a ajustes globales (Houtman *et al.* 2007) del modelo de unión 2:1 a los datos de titulaciones directas y reversas para cada proteína homóloga. Los parámetros de los ajustes individuales y globales al modelo de unión 2:1, se encuentran representados en la **Tabla 1**. Los experimentos fueron realizados en el tampón químico 200 mM de fosfato de sodio pH 7, 0.2 mM de DTT a 25°C.

Tabla V.1. Parámetros de ajustes individuales y globales en la interacción E2C:DBS.

Proteína	ΔH_{coop} (kcal/mol)	K_{coop}	ΔG_{coop} (kcal/mol)	$-T \cdot \Delta S_{\text{coop}}$ (kcal/mol)	N^a	$\chi^2{}^b$	$\chi^2_{K_{\text{coop}}=1}{}^b$
E2C-16 _{Directa}	0.97±1.07	2.72±0.10	-0.59±0.10	-1.57±1.05	0.4±0.1	97	132
E2C-16 _{Reversa}	0.44±1.06	2.35±0.10	-0.51±0.10	-0.94±1.10	1.6±0.1	83	107
E2C-16 _{Reversa} ^c	1.35±1.60	1.95±0.11	-0.40±0.15	-1.74±1.61	1.8±0.1	71	131
E2C-16 _{Global} ^d	0.89±0.48	2.60±0.32	-0.57±0.07	-1.46±0.49		56	109
E2C-16 _{Global} ^{cd}	0.88±0.48	2.34±0.29	-0.50±0.07	-1.38±0.49		79	173
E2C-11 _{Directa}	-1.14±0.86	3.45±1.11	-0.73±0.19	0.41±0.88	0.3±0.2	10	11
E2C-11 _{Reversa}	-3.56±0.40	3.30±0.55	-0.71±0.10	2.85±0.41	1.9±0.1	20	21
E2C-11 _{Global} ^d	-3.07±0.53	3.42±0.78	-0.73±0.14	2.34±0.55		27	42
E2C-18 _{Directa}	-0.40±1.39	0.58±0.12	0.32±0.12	0.72±1.40	0.5±0.1	78	70
E2C-18 _{Reversa}	0.01±0.01	1.23±0.19	-0.12±0.11	0.13±0.11	1.3±0.1	121	185
E2C-18 _{Global} ^d	2.25±1.14	0.80±0.14	0.14±0.11	-2.12±1.15		163	166

^a Las estequiometrías de la reacción para las titulaciones directas y reversas fueron calculadas a partir de la fracción incompetente del ligando (Houtman *et al.* 2007). ^b Valores de χ^2 de los ajustes a un modelo de dos sitios y a un modelo de dos sitios independientes. ^c Parámetros termodinámicos de una titulación reversa adicional para E2C-16 y su ajuste global correspondiente (ver **Figura 5**). ^d Ajustes globales a un modelo de dos sitios de titulaciones directas y reversas (ver **Figura 4A-4B**, **4C-4D** y **4E-4F**). Los errores provienen de simulaciones de Monte-Carlo para cada ajuste de datos (Houtman *et al.* 2007).

Las estequiometrías de los ajustes son del orden esperado, coincidiendo con las estequiometrías directas y reversas seguidas por fluorescencia y EMSA (**Figura 2**). Los parámetros obtenidos para los ajustes individuales coinciden con los de los ajustes globales, sus errores disminuyen considerablemente al ajustar globalmente (**Tabla 1**). Por la robustez de los ajustes globales (Beechem 1992), de ahora en más nos centraremos en el análisis de los parámetros resultantes de los ajustes globales al modelo de dos sitios de los datos de titulaciones ‘directas’ y ‘reversas’.

El ajuste global de los datos de la formación del complejo ternario E2C-16:DBS:E2C-16 (**Figura 4A-4B**) indica que E2C-16 reconoce al sitio en tándem propio con cooperatividad positiva, $K_{\text{coop}} = 2.6 \pm 0.3$ (**Tabla 1**) ($\Delta G_{\text{coop}} = -0.6 \pm 0.1$ kcal/mol). Como control, el ajuste global de un modelo de dos sitios independientes (fijando K_{coop} en 1) a los datos, resultó en un incremento en dos veces del valor de χ^2 del ajuste (**Tabla 1**) confirmando que la formación del complejo ternario es cooperativa en E2C-16. Adicionalmente, el ajuste global con otros datos independientes coincide, $K_{\text{coop}} = 2.3 \pm 0.3$, indicando que el error experimental es del orden del error del ajuste (**Figura 5** y **Tabla 1**). La disección del ΔG_{coop} en sus componentes entálpicos y entrópicos se muestra en la **Figura 6**. Sorprendentemente, para E2C-16 la cooperatividad positiva se encuentra gobernada por la entropía, con un valor de $-T\Delta S_{\text{coop}} = 1.5 \pm 0.5$ kcal/mol (**Tabla 1**).

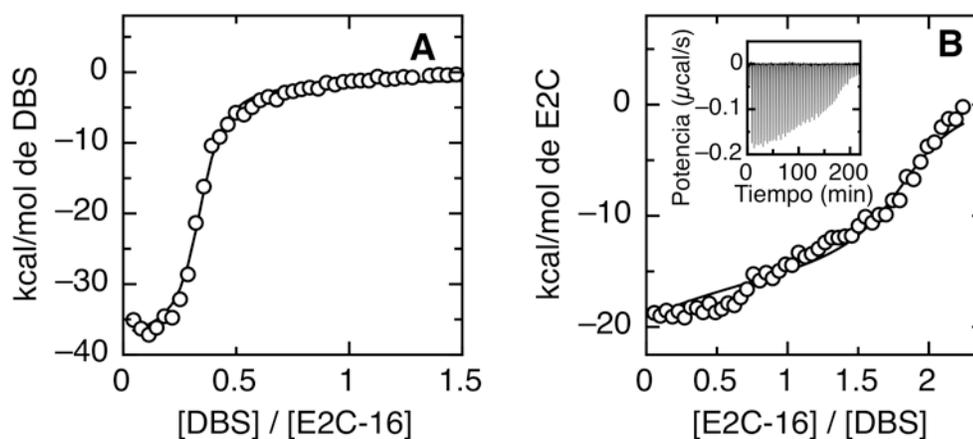


Figura V.5. Ajuste global adicional de E2C-16. Las líneas corresponden al ajuste global (Houtman *et al.* 2007) entre la titulación directa panel (A) (ver **Figura 4A**) y la titulación reversa adicional panel (B). Los parámetros de ajuste se encuentran representados en la **Tabla 1**. Los experimentos fueron realizados en el tampón químico 200 mM de fosfato de sodio pH 7, 0.2 mM de DTT a 25°C.

La interacción entre las proteínas homólogas E2C-11 y E2C-18 y el DBS correspondiente al genoma de HPV16 (**Figura 1D**), presentó un panorama llamativamente distinto al de E2C-16 (**Figura 4**, **Tabla 1**). Por un lado, el reconocimiento de la proteína E2C-11 al DBS mostró una constante de afinidad adicional, K_{coop} de 3.4 ± 0.8 ($\Delta G_{\text{coop}} = -0.7 \pm 0.1$), y el ajuste empeora considerablemente al fijar la K_{coop} en 1, un resultado similar al de E2C-16 (**Tabla 1**). Por otro lado, la proteína E2C-18 se une a los sitios dispuestos en tándem de manera no-cooperativa con una K_{coop} de 0.8 ± 0.2 (ΔG_{coop} de 0.1 ± 0.1 kcal/mol). El valor de χ^2 se incrementa en forma mínima al incluir el término de cooperatividad en el modelo de unión (**Tabla 1**) de acuerdo con un

reconocimiento meramente aditivo. A modo ilustrativo, representamos las consecuencias de la unión al DBS para los tres casos simulando la población fraccional de DBS libre, DBS con un sitio ocupado y DBS con ambos sitios ocupados por las distintas proteínas homólogas (**Figura 6**, paneles de abajo). Las líneas sólidas fueron calculadas usando los valores medidos de K_{BS1} , K_{BS2} , y K_{coop} y las líneas punteadas fueron calculadas usando los valores medidos de K_{BS1} , K_{BS2} fijando K_{coop} en 1 (ver materiales y métodos). La unión con cooperatividad positiva para E2C-16 y E2C-11 se manifiesta con dos efectos en simultáneo: en forma relativa a la unión no cooperativa, el complejo ternario se puebla a menores concentraciones de E2C y el complejo binario se puebla en menor grado (**Figura 6B** y **6D**). Para E2C-18 no se observan diferencias, como era esperable de acuerdo con la falta de cooperatividad (**Figura 6E**).

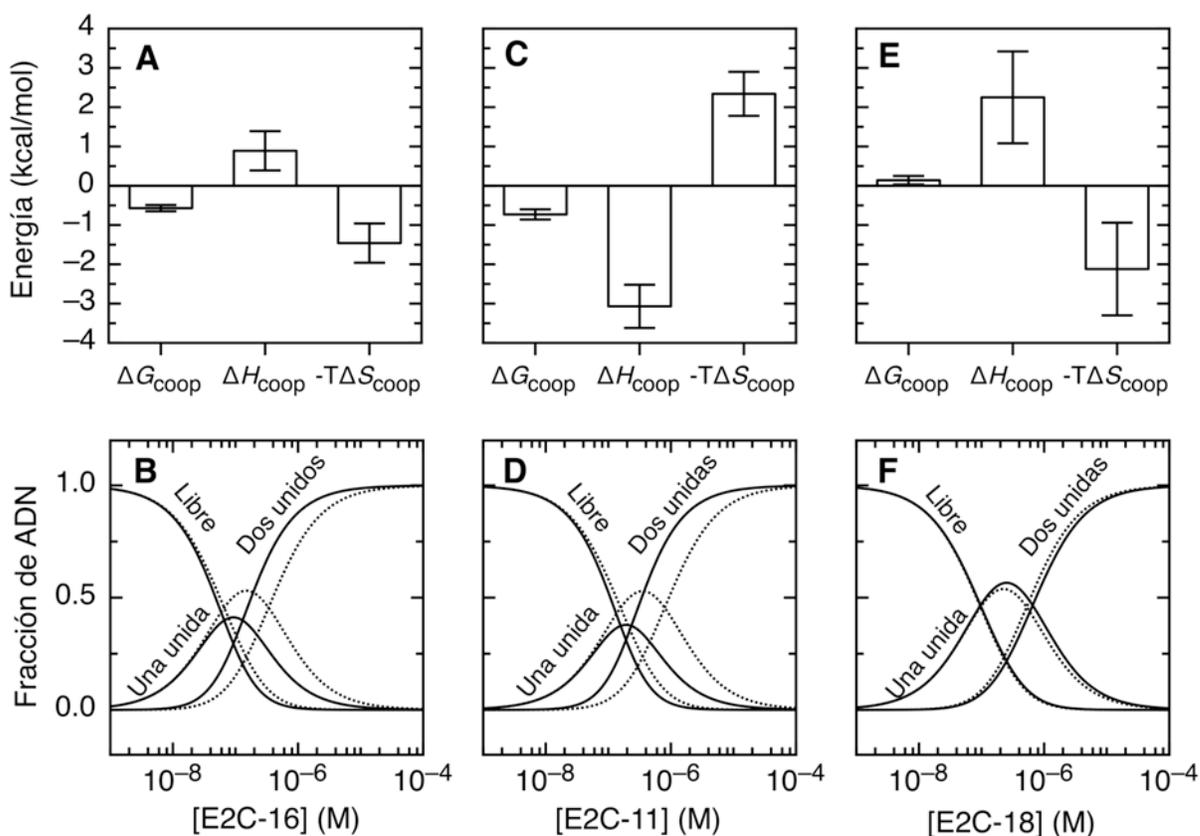


Figura V.6. Unión cooperativa de E2C al sitio tándem DBS. Paneles de arriba: disección termodinámica de la unión de (A) E2C-16, (C) E2C-11 (E), E2C-18 a DBS. Valores negativos representan un cambio favorable adicional en la contribución a la unión, mientras que valores positivos representan una contribución adicional desfavorable para la unión de DBS. Paneles de abajo: población fraccional de DBS libre, con un sitio ocupado (una unida) y DBS con los dos sitios ocupados (dos unidas) en función de la concentración de (B) E2C-16, (D) E2C-11 y (F) E2C-18. Las líneas sólidas fueron calculadas usando los valores medidos de K_{BS1} , K_{BS2} , y K_{coop} . Las líneas punteadas fueron calculadas usando los valores medidos de K_{BS1} , K_{BS2} , y $K_{coop}=1$. Los experimentos fueron realizados en el tampón químico 200 mM de fosfato de sodio pH 7, 0.2 mM de DTT a 25°C.

Al analizar la disección termodinámica en conjunto, las diferencias entre proteínas homólogas para el reconocimiento del doble sitio DBS resaltan aún más marcadas. Efectivamente, en contraste al modo de unión de E2C-16, la cooperatividad positiva de E2C-11 posee un origen entálpico, con un ΔH_{coop} de -3.1 ± 0.5 kcal/mol (**Figura 6C** y **Tabla 1**), siendo la unión entrópicamente desfavorable. Es llamativo que esta diferencia en el origen energético para la unión sinérgica de E2C-16 y E2C-11 al DBS no se aprecia en las titulaciones vistas para los sitios aislados BS1 y BS2 que conforman el DBS (**Figura IV.6** y **Tabla IV.4**) (ver Discusión). En el caso de E2C-18, el valor de casi cero de ΔG_{coop} se origina en la compensación de contribuciones entálpicas y entrópicas, las cuales son de magnitud similar a las observadas para E2C-11 y E2C-16 (**Figura 6E**). Esto indica que la falta de sinergismo en E2C-18 es producto de compensaciones termodinámicas, y por lo tanto la unión de la segunda molécula de E2C-18 al DBS no es estrictamente ‘invisible’ frente a la unión de la primera molécula al DBS. Con el fin de investigar desde otra perspectiva la formación del complejo ternario, analizamos el comportamiento estructural del DBS frente la unión de E2C. Además de buscar profundizar los resultados obtenidos por ITC, siendo que el DBS constituye en simultáneo el *ori* y la región promotora de HPV, resulta clave analizar la plasticidad conformacional del ADN inducida por las moléculas que lo reconocen específicamente (Alexandrov *et al.* 2010).

Cambios conformacionales en la interacción entre E2C y DBS.

Utilizamos espectroscopía de CD y absorbancia como reporteras del cambio en la conformación del ADN de DBS luego de la formación del complejo ternario con E2C, y comparamos con los cambios asociados con la unión de los sitios individuales BS1 y BS2 (**Figuras 7** y **8**). En la **Figura 7A** y **7B** se representan los espectros correspondientes al sitio BS1 libre (líneas discontinuas) y al complejo binario formado con E2C-16 (rojo), E2C-11 (azul) o E2C-18 (verde). En la **Figura 7C** y **7D** se representa lo mismo pero para el sitio BS2, analizado el Capítulo anterior, **Figura IV.5**. La formación del complejo entre E2C y BS1 presenta cambios conformacionales en el BS1, al igual que para el caso del BS2, manteniendo la estructura nucleotídica de forma-B. Para ambos oligonucleóticos, las diferencias de señal espectral entre las tres proteínas homólogas es mínima, indicando una deformación del ADN similar en ambos complejos aislados.

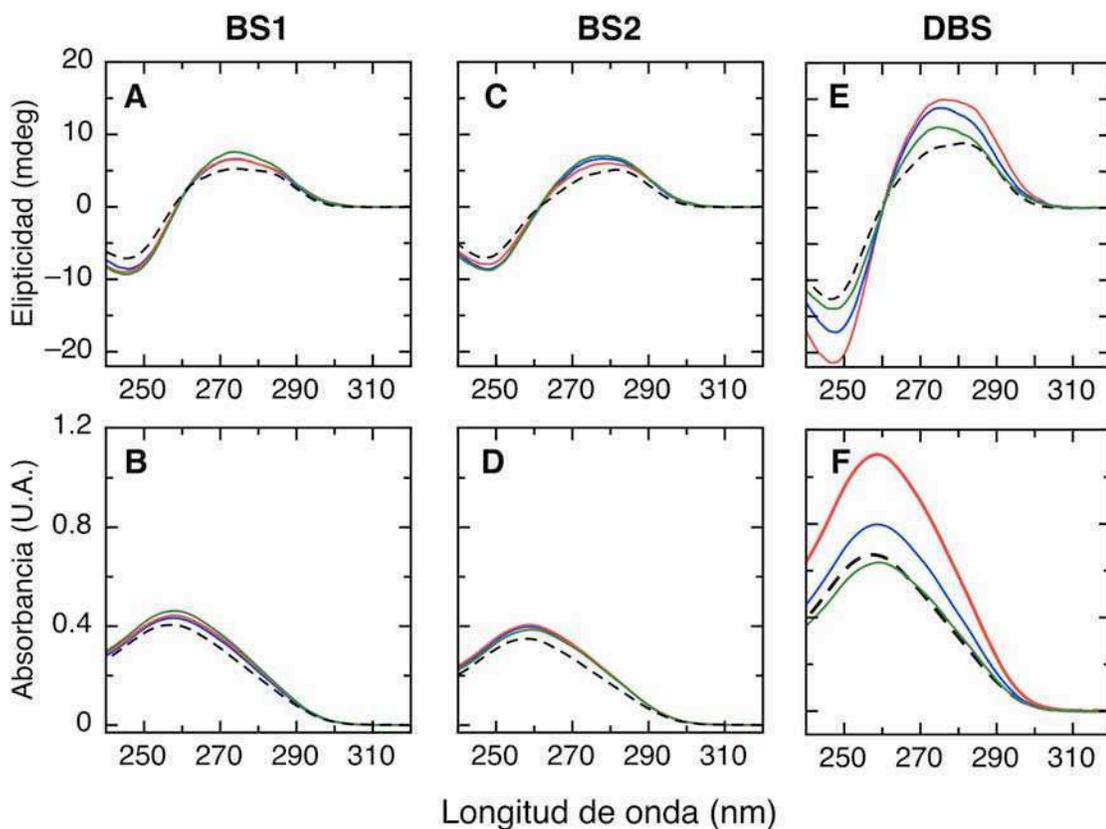


Figura V.7. Espectros de CD y absorbancia de los complejos binarios y ternarios y ADN libre. Cada panel muestra el espectro de CD o absorbancia de ADN libre (líneas discontinuas) y unido (líneas continuas) a E2C-16 (rojo), E2C-11 (azul), E2C-18 (verde) para BS1 (**A** y **B**), BS2 (**C** y **D**) y DBS (**E** y **F**). Los experimentos fueron realizados en el tampón químico 200 mM de fosfato de sodio pH 7, 0.2 mM de DTT a 25°C. Las especies libres y unidas fueron medidas a una concentración de 15 μ M en una estequiometría de 1:1 para los complejos binarios y de 2:1 para los complejos ternarios. A los espectros de complejos mostrados se les restaron los espectros de la proteína libre correspondiente.

Posteriormente registramos los espectros correspondientes al ADN DBS aislado y al complejo ternario E2C:DBS:E2C. En primer lugar analizamos los espectros crudos para las tres proteínas homólogas (**Figura 7E** y **7F**). El espectro de CD de DBS libre (**Figura 7E**, líneas discontinuas) muestra un pico positivo a 275 nm, un cruce de la señal en cero a 260 nm y un pico negativo a 245 nm. Dichas características son propias de ADN estructurado en forma-B (Baase y Johnson 1979; Gray *et al.* 1995; Gray 1996). Por lo tanto, el sitio BS1 y BS2 dispuesto en tándem, DBS, mantiene la estructura secundaria de forma-B característica de cada sitio individual, ver **Figuras IV.5**, **IV.7A** y **IV.7C** y (Kim *et al.* 2000; Hizver *et al.* 2001; Hooley *et al.* 2006). Al formar el complejo ternario con E2C-16, E2C-11 o E2C-18 (**Figura 7E** y **7F** respectivamente, líneas continuas), el sitio DBS mantiene su estructura secundaria de ADN-B. Además, podemos observar para las tres proteínas cambios en la intensidad de la señal del pico

positivo a 275 nm y del pico negativo a 245 nm. Dichos cambios en el DBS unido varían notablemente entre las proteínas homólogas. Con el fin de discriminar si los cambios conformacionales se encuentran asociados a la unión cooperativa o a la suma de los cambios estructurales de los complejos binarios, a continuación comparamos los espectros diferenciales.

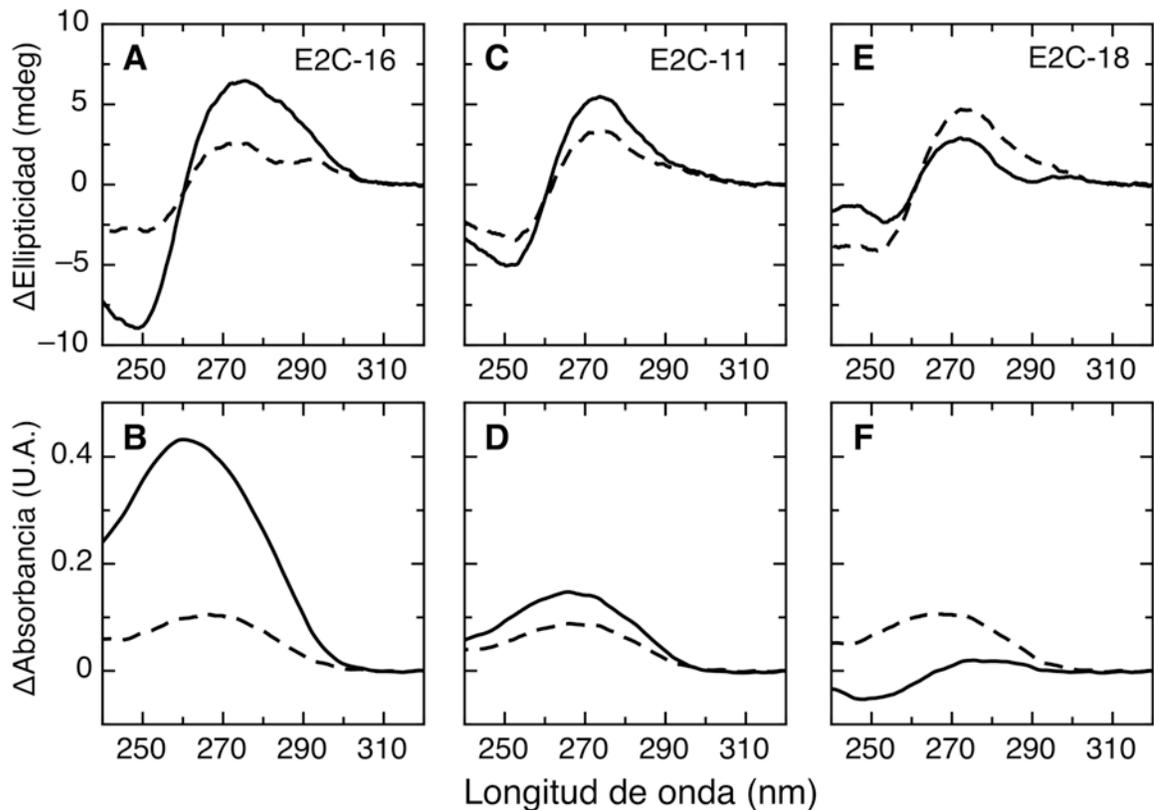


Figura V.8. Cambios espectroscópicos asociados con la unión de E2C al sitio en tándem DBS. Líneas sólidas: espectro del complejo ternario E2C:DBS:E2C. Líneas discontinuas: suma del espectro diferencial para los complejos binarios E2C:BS1 y E2C:BS2. Paneles de arriba: espectros diferenciales de absorbanza para (A) E2C-16, (C) E2C-11 y (E) E2C-18. Paneles de abajo: espectros diferenciales de CD para (B) E2C-16, (D) E2C-11 y (F) E2C-18. Los experimentos fueron realizados en el tampón químico 200 mM de fosfato de sodio pH 7, 0.2 mM de DTT a 25°C.

Obtuvimos los espectros diferenciales de CD y absorbanza para el complejo ternario restando el espectro del oligonucleótido DBS libre y el de E2C (en una relación molar 1:2) del espectro del complejo E2C:DBS:E2C (**Figura 8**). La combinación de espectros diferenciales de CD y absorbanza representan los cambios en la estructura del ADN DBS al unirse a E2C (**Figura 8**, líneas continuas). Luego comparamos el espectro diferencial (ambos CD y absorbanza) para el complejo ternario con la suma del espectro diferencial para cada complejo binario individual de E2C:BS1 y E2C:BS2 (**Figura 8**, líneas discontinuas). Llamativamente, hay diferencias evidentes entre ambos espectros para todas las proteínas, indicando que la formación

de complejo ternario involucra un reordenamiento conformacional en el ADN DBS que va más allá del observado para los sitios simples (**Figura 7**). Notablemente, la diferencia entre la conformación del DBS (líneas continuas) y la suma de los sitios individuales BS1 y BS2 (líneas discontinuas) varía entre los tres dominios. Esto sugiere que los cambios estructurales en el DBS asociados con la formación del complejo ternario dependen principalmente del dominio de E2C y no de la secuencia de bases. Lo más llamativo corresponde al caso de E2C-18, dónde se pueden observar cambios estructurales en el DBS a pesar de unirse al mismo en forma aditiva.

Cambios estructurales del DBS en función del tiempo.

Finalmente, investigamos la cinética de la formación de los complejos ternarios de E2C con DBS utilizando la técnica de flujo-detenido (ver materiales y métodos). Para ello, mezclamos volúmenes idénticos de 2 μM de E2C-16 y 1 μM de DBS y seguimos los cambios de la absorbancia. El evento de asociación entre E2C y ADN determinado en nuestro laboratorio así como por otros laboratorios, mostró una constante de asociación de 0.1 a $1.4 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$ (Ferreiro y Prat Gay 2003; Hooley *et al.* 2006). Esto implica que en las condiciones de medición presentes, que requieren de altas concentraciones de reactantes para obtener buena relación señal/ruido, las dos moléculas se asocian dentro del tiempo muerto del equipo (aproximadamente 2 ms), y cualquier cambio en la señal observada corresponde a reordenamientos conformacionales unimoleculares. Para la proteína E2C-16 y bajo estas condiciones de medición, los cambios en la señal de absorbancia observados al equilibrio (**Figuras 7 y 8**) ocurren dentro del tiempo muerto del equipo (datos no mostrados). Por lo tanto, en el caso de E2C-16, el reordenamiento conformacional en el DBS ocurre en la escala de tiempo del sub-milisegundo, en paralelo con el evento de asociación bimolecular. Los cambios en la absorbancia correspondientes a la unión de E2C-18 al DBS ocurren dentro del tiempo muerto del equipo, sugiriendo que los cambios conformacionales al unir al ADN y la asociación ocurren también en forma concertada.

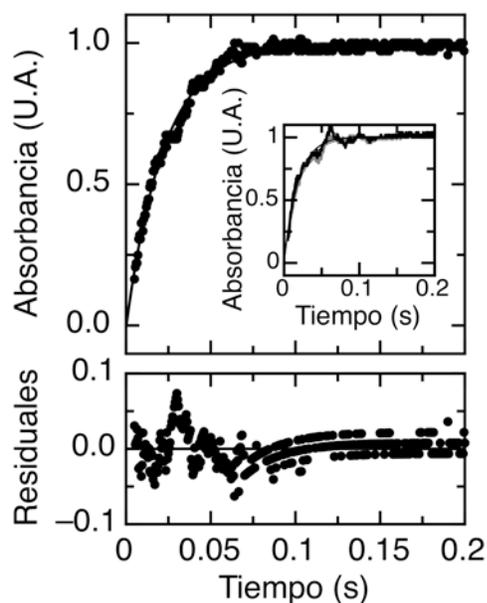


Figura V.9. Cinética del cambio estructural del sitio de ADN en tándem DBS al unir E2C. 1 μM de DBS fue mezclado con 2 μM de E2C-11 en 200 mM de fosfato de sodio pH 7, 0.2 mM de DTT a 25°C y se siguieron los cambios en absorbancia a 270 nm. Se representan los datos normalizados y el ajuste a una función exponencial. En el recuadro se representa la mezcla de 1 μM de DBS con 2 μM y 8 μM de E2C-11, cuyas constantes de velocidad resultantes del ajuste a una exponencial simple corresponden a $50.1 \pm 0.9 \text{ s}^{-1}$ y $54.7 \pm 1.2 \text{ s}^{-1}$ respectivamente.

A diferencia de E2C-16 y E2C-18, los cambios en la conformación del ADN DBS al unir a E2C-11 fueron observados dentro de la ventana experimental (**Figura 9**). La cinética se describe adecuadamente con el ajuste a una función exponencial simple de constante de velocidad de $44.9 \pm 0.5 \text{ s}^{-1}$. Esta constante de velocidad es independiente de la concentración de E2C en el rango de 2 a 8 μM (**Figura 9**, recuadro), indicando que el paso limitante de la reacción bajo estas condiciones es un proceso unimolecular. Esto implica que para la proteína E2C-11, el complejo ternario conduce a una transición conformacional en el DBS luego del evento de asociación, el cual basado en las constantes de asociación conocidas ocurre dentro del tiempo muerto del equipo (Ferreiro y Prat Gay 2003; Hooley *et al.* 2006).

Cuarta Sección

Discusión

'Would you tell me, please, which way I ought to go from here?'

'That depends a good deal on where you want to get to,' said the Cat.

'I don't much care where,' said Alice.

'Then it doesn't matter which way you go,' said the Cat.

'But I don't want to go among mad people,' Alice remarked.

'Oh, you can't help that,' said the Cat: 'we're all mad here. I'm mad. You're mad.'

'How do you know I'm mad?' said Alice.

'You must be,' said the Cat, 'or you wouldn't have come here.'

Lewis Carroll, Alice's Adventures in Wonderland

En el presente trabajo de tesis estudiamos diferentes aspectos del comportamiento de la proteína E2C mediante estudios en solución y mediante análisis de secuencia. En primer lugar experimentamos con la plasticidad conformacional de la molécula E2C monomerizada frente al aumento de la temperatura y describimos el mecanismo de formación de protofibras amiloides y el paisaje morfológico (Capítulo I). En segundo lugar investigamos la discriminación de secuencias específicas de ADN por E2C mediante un análisis estadístico (Capítulo II) y mediante calorimetría de titulación isotérmica (Capítulo III). Luego, analizamos la influencia de la divergencia de secuencia de E2C en la estabilidad conformacional del dominio y en la interacción con ADN (Capítulo IV). Finalmente, abordamos la termodinámica de la unión cooperativa de E2C a dos sitios de ADN adyacentes (Capítulo V). Globalmente, utilizamos como modelo de estudio al dominio C-terminal E2 para intentar comprender su rol en las finas regulaciones del ciclo infeccioso de papilomavirus, como ser la capacidad de discriminación de secuencias específicas. También, para profundizar el conocimiento de aspectos generales de interacciones entre proteínas y ADN, como ser el reconocimiento de sitios en tándem y mecanismos de lectura indirecta. Y finalmente para comprender aspectos universales en la formación de agregados estructurados causantes de múltiples enfermedades, principalmente neurodegenerativas. A continuación iniciaremos la discusión, donde buscaremos contextualizar los resultados mediante datos bibliográficos, profundizar los resultados mediante figuras integradoras y generar nuevos interrogantes mediante resultados parciales no mostrados en la sección resultados.

Plasticidad conformacional de E2C.

A partir de los resultados expuestos en el Capítulo I, construimos en la misma escala dimensional un esquema tentativo del mecanismo de polimerización de scE2C integrando los estudios cinéticos y la caracterización morfológica (**Figura 1**). La polimerización de scE2C produce especies con características de tipo amiloide (**Figura I.2**), y se desencadena a partir de la formación de un núcleo estructurado compuesto por 2 ± 1 monómero de scE2C, el cual se mantiene estable a lo largo de la reacción de polimerización (**Figura I.7**). El análisis morfológico por criomicroscopía muestra la presencia de especies que coinciden con la molecularidad del núcleo (**Figura I.10**), reforzando el modelo cinético del mecanismo de polimerización de scE2C. La fase de elongación comprende la sucesiva adición de dos moléculas de scE2C (**Figura I.7**),

dando lugar a un paisaje de ensamblados oligoméricos definidos de tipo protofibrilar (**Figura I.5**). A partir del análisis minucioso del diámetro hidrodinámico de las especies medidas por DLS (**Figura I.3** y **Tabla I.1**) y del análisis estadístico del volumen de las especies observadas por AFM (**Figura I.5** y **Tabla I.2**), definimos que las bolas pequeñas están compuestas por 13 ± 6 unidades de scE2C, las bolas mayores están a su vez compuestas por 15 ± 3 bolas pequeñas y finalmente los gusanos están compuestos por entre 2 a 15 bolas mayores esparcidas tipo collar de perlas (**Figura 1**). Vale aclarar que para las estimaciones utilizamos volúmenes medidos y calculados, lo cual implica la asunción de morfologías de geometría ‘blanda’ y no rígida. Apoyando esta idea, la unión significativa de ANS y de la Tioflavina T observada (**Figura I.2**), sugieren cierta flexibilidad, accesibilidad de moléculas de agua al interior de la proteína, en las especies oligoméricas encontradas (Jha y Udgaonkar 2009; Bolognesi *et al.* 2010). El detalle y la calidad de los datos medidos generados a partir del análisis integral de un mismo proceso de polimerización por diferentes técnicas, nos indujo a conjeturar sobre la presencia de una jerarquía oligomérica dentro del paisaje morfológico observado, la cual representamos en la **Figura 1**.

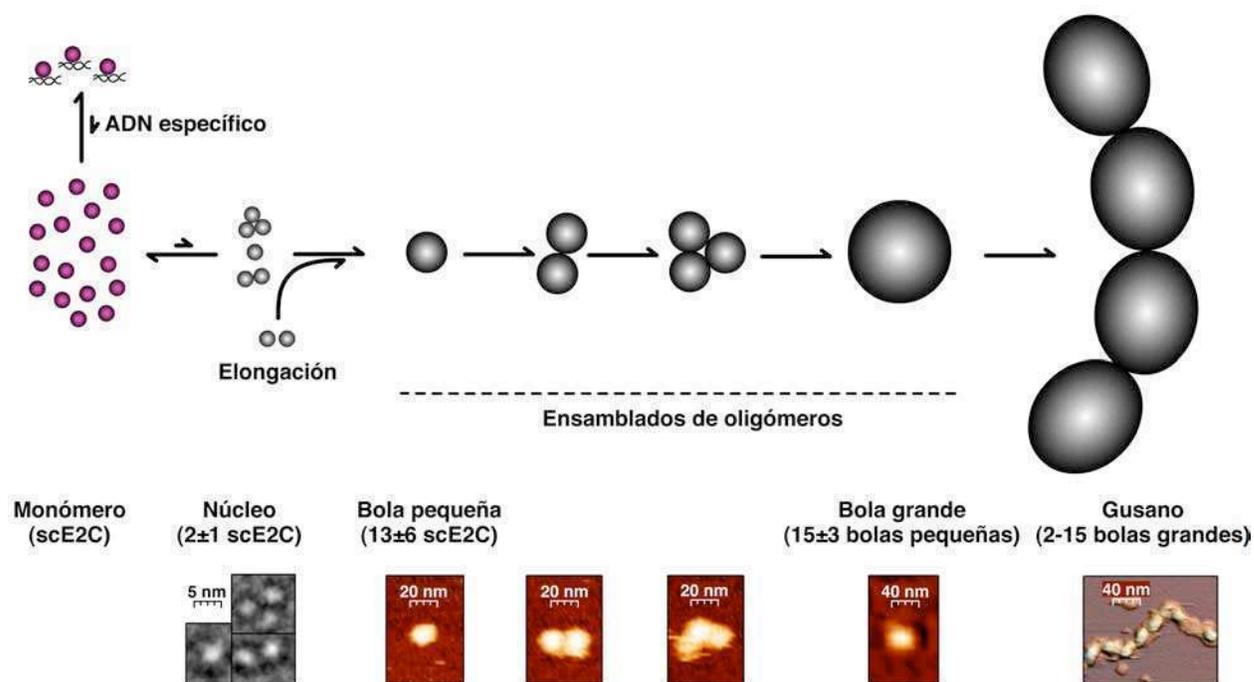


Figura 1. Formación jerárquica de intermediarios amiloideos en la vía de polimerización de scE2C. Esquema integrador construido a escala mediante los resultados expuestos en el Capítulo I. El esquema contempla los resultados de AFM, DLS, cinética y microscopía electrónica. Abajo ejemplificamos las diferentes especies observadas con su respectiva escala. Creado en Flash v 8.0, Macromedia.

De acuerdo con nuestra hipótesis jerárquica, completamos el paisaje morfológico de la agregación ordenada de scE2C añadiendo dos intermediarios del pasaje de bolas pequeñas a bolas mayores. Dichos intermediarios corresponden a ensamblados oligoméricos compuestos por dos y tres bolas pequeñas. No tuvimos en cuenta estos intermediarios en el análisis morfológico expuesto en la sección resultados (**Figura II.5**), dado que en todas las imágenes de AFM procesadas sólo pudimos observar no más de cinco ejemplares para cada especie. Esta observación sugiere que dichos dímeros y trímeros de bolas pequeñas son inestables en relación con las bolas pequeñas y mayores aisladas. Estos ‘eslabones’ estarían revelando que la progresión de bolas pequeñas a bolas mayores posee una baja barrera energética de transición, reforzando la direccionalidad en nuestra hipótesis jerárquica para la formación de las unidades elementales de la morfología tipo gusano. En forma similar, las bolas mayores también fueron vistas en baja proporción en relación a las bolas pequeñas (**Tabla I.2**), indicando también inestabilidad y justificando la ‘predisposición’ a formar gusanos tipo collar de perlas. Finalmente, la correlación ‘Fandrich’ entre los parámetros cinéticos $T_{50\%}$ y VE (Fandrich 2007) (**Figura I.8**), indica una estricta asociación entre las fases de nucleación y elongación en el proceso de polimerización de scE2C, reforzando en cierto modo la progresión jerárquica en el paisaje morfológico observado, donde la especie oligomérica más abundante corresponde a las bolas pequeñas, las cuales al aumentar su población interactúan entre sí dando lugar a las especies de mayor tamaño.

Comenzamos el Capítulo I con una breve reseña acerca de los múltiples caminos de agregación de E2 observados en distintas condiciones y frente a distintos estímulos (Ferreiro *et al.* 2000; Wetzler *et al.* 2007; Smal *et al.* 2009). De acuerdo a esto, es posible separar globalmente en dos los estados agregados de E2: estados agregados desordenados (Ferreiro *et al.* 2000; Smal *et al.* 2009) y estados agregados ordenados ((Wetzler *et al.* 2007) y Capítulo I), los cuales como apreciamos en la **Figura I.1**, se mezclan en E2C aunque no su versión monomerizada. Además, dentro de los estados ordenados observamos tanto estructuras flexibles protofibrilares (**Figuras I.4, I.5 y I.9**), como estructuras fibrilares rígidas (Wetzler *et al.* 2007). En base a esto nos preguntamos primero: ¿Los estados ordenados y desordenados forman parte de un mismo camino o proceso? Y segundo: ¿Las protofibras observadas en las **Figuras I.4, I.5, I.9** y (Wetzler *et al.* 2007) son realmente intermediarios de las fibras rígidas observadas mediante la inducción por TFE (Wetzler *et al.* 2007)? ¿Forman parte del mismo camino o son un camino

independiente?

En primer lugar, como mencionamos anteriormente, las interacciones E2C-ADN y E2C-E7 fomentan la agregación desordenada de E2C, y en ambos casos la región estructural de E2C involucrada en la interacción corresponde a la hélice $\alpha 1$ (Hegde y Androphy 1998; Smal *et al.* 2009). En efecto, no solo la formación de homo-oligómeros ordenados descrita en el Capítulo I es inhibida por la presencia de ADN específico (**Figura I.1** y (Wetzler *et al.* 2007)) sino que también impide la formación de co-oligómeros ordenados con la proteína viral E7 (Smal *et al.* 2009). Por lo tanto, en principio, la región involucrada en el camino hacia hetero-agregados desordenados se encontraría en otra región de la proteína E2C que la hélice $\alpha 1$. Entonces, el hecho de estar la hélice $\alpha 1$ ocupando el surco mayor del ADN específico es suficiente para evitar la oligomerización estructurada de E2C, dando un protagonismo clave a esta región de E2C en este proceso. Tomando este análisis en conjunto, se descarta la posibilidad que los caminos hacia estados desordenados y ordenados formen parte de una misma ruta y se puede aseverar que el camino hacia complejos ordenados es específico de secuencia. Efectivamente, el comportamiento diferencial tanto frente a la desnaturalización térmica (ver datos parciales en la **Figura 5**) como a la propensión a formar agregados amiloideos de los distintos dominios homólogos de E2C estudiados en el Capítulo III (datos parciales no mostrados), garantizan la especificidad de secuencia en la propensión amiloidea de E2C.

En segundo lugar nos preguntamos si el camino hacia fibras rígidas es el mismo que el camino hacia protofibras de tipo gusano y ensamblados oligoméricos. Esta pregunta es aún más compleja de estudiar que la anterior, ya que en la literatura se encuentran proteínas que forman fibras sin la necesidad de intermediarios esféricos o protofibrilares tanto como de proteínas que a veces los usan de intermediarios (Kodali y Wetzel 2007). Por lo tanto a nivel general aún no hay una respuesta simple. En nuestro caso, utilizamos la monomerización de scE2C como herramienta de estudio de estos intermediarios, aunque no queda del todo claro aún si éstos forman parte de mismo camino hacia las fibras rígidas observadas únicamente en la versión natural dimérica (Wetzler *et al.* 2007). Por lo tanto, como nuestro abordaje se focalizó en el estudio de los intermediarios amiloideos, no podemos concluir si ambos caminos compiten entre si o no, quedando abierto el interrogante. Otro aspecto a discutir consiste en entender por qué la monomerización de E2C permitió estudiar en profundo detalle dichos intermediarios y qué ventaja evolutiva le confiere al dímero esta diferencia frente al monómero diseñado

artificialmente.

Muchas proteínas no involucradas en enfermedades amiloideas han sido utilizadas como herramientas para el estudio de mecanismos de polimerización de estados proteicos agregados en general. El caso más estudiado corresponde a la proteína pequeña Barstar, a partir de la cual se logró describir en detalle el mecanismo de formación amiloidea (Kodali y Wetzel 2007), la presencia de caminos de agregación paralelos (Kumar *et al.* 2009; Kumar *et al.* 2009) y la heterogeneidad conformacional de las protofibras (Sekhar *et al.* 2011). Por otro lado, la acumulación patológica de agregados amiloideos ha sido asociada con varios tipos de cáncer como es el caso de la proteína p53 (Ishimaru *et al.* 2003). Este caso en particular nos interesa dado que p53 une a ADN y curiosamente, su unión a ADN específico inhibe el camino hacia la agregación ordenada (Ishimaru *et al.* 2009; Silva *et al.* 2010) pero no así el inespecífico (Ishimaru *et al.* 2009). En papilomavirus, la acumulación de la proteína E7 en estado oligomérico con características amiloideas se observó en tejidos de células cáncer de cuello de útero (Dantur *et al.* 2009). Como mencionamos anteriormente, la interacción E7-E2C facilita la formación de agregados oligoméricos (Smal *et al.* 2009). Frente a este panorama no podemos descartar que los agregados de E2C sean como Barstar, únicamente un caso ideal para estudiar agregados en general, sino que al contrario hay evidencias de posibles implicancias en amiloidosis y cáncer. Más allá de esta hipótesis, la idea de utilizar la monomerización de E2C como herramienta para estabilizar el estado intermediario y estudiarlo en profundidad, es para nuestro conocimiento el primer ejemplo en la literatura. Resta preguntarse entonces qué efecto en la polimerización de E2C genera presencia de ADN inespecífico, pues también la respuesta a esta pregunta puede tener implicancias biológicas en el virus.

Finalmente, llama la atención que la proteína EBNA cuya estructura cristalográfica y topología es cuasi idéntica a la de E2C, tanto en estado libre como unida a ADN (Bochkarev *et al.* 1995; Bochkarev *et al.* 1996), también presenta la formación de oligómeros solubles esféricos (Freire *et al.* 2008). Este paralelismo entre estos dos virus alejados entre sí nos induce a reflexionar si los agregados amiloideos de tipo intermediario poseen una función determinada en el virus papiloma. A medida que se estudian más casos, aumenta la variedad de funcionalidades posibles asociadas con agregados amiloideos protofibrilares (por ejemplo, (Maji *et al.* 2009)), a diferencia de las inertes fibras rígidas (Kodali y Wetzel 2007). Además, la arquitectura beta cruzada propia de fibras rígidas se encuentra también presente en las protofibras flexibles

funcionales (Fowler *et al.* 2007), sugiriendo que la funcionalidad rige mayoritariamente en la morfología en lugar de la secuencia que define el estado amiloideo. El rol que juega el ADN específico inhibiendo la agregación ordenada de E2C sería entonces clave en la funcionalidad de dichos agregados, más aún si esto no ocurre con el ADN inespecífico, la principal regulación de la vía amiloidea de E2 podría estar dictada por el ADN. En esta línea de razonamiento, los niveles de expresión de E2 en las células infectadas pueden estar vinculados con la ruta amiloidea además de como mencionamos anteriormente y discutiremos a continuación, estar vinculado con el desarrollo de cáncer uterino.

Discriminación de secuencias específicas de ADN por la proteína E2.

La interacción E2C-ADN conforma un complejo altamente dinámico, con una interfaz húmeda (Hegde 2002; Ferreiro *et al.* 2005) en la cual las contribuciones energéticas de los residuos de la proteína son aditivos según se describió en nuestro laboratorio (Ferreiro *et al.* 2005). De acuerdo con esto, fuimos capaces de describir la evolución del sitio de ADN utilizando un modelo que postula contribuciones aditivas en la energía asignada a las bases del ADN involucradas en la unión (**Figura II.4**). Hemos propuesto que la interfaz flexible del complejo E2-ADN permite al ADN evolucionar mucho más rápido que la proteína y en cuatro sitios de ADN bien diferenciados (**Figura II.2**). Las cuatro bases ubicadas en el espaciador del sitio de unión de E2 son responsables de la lectura indirecta y se encuentran significativamente conservadas (**Figuras II.2 y II.3**), sugiriendo que su contribución energética juega un rol en la evolución de la función génica de papilomavirus humano. La conservación del gradiente dentro del espaciador (**Figuras II.2 y II.3**) contradice la creencia general según la cual las cuatro bases realizan aproximadamente contribuciones igualitarias a la unión (Bedrosian y Bastia 1990; Hegde y Androphy 1998; Hines *et al.* 1998; Rozenberg *et al.* 1998; Kim *et al.* 2000; Hizver *et al.* 2001; Hegde 2002; Dell *et al.* 2003; Zimmerman y Maher 2003; Byun y Beveridge 2004; Zhang *et al.* 2004; Rohs *et al.* 2005; Blakaj *et al.* 2006), poniendo en evidencia que la lectura indirecta (**Figura 4 Introducción**) puede generar patrones complejos de conservación.

Los sitios de unión de E2 más regulares no son palíndromos perfectos, en contraste con la naturaleza homodimérica del dominio de E2C. Los sitios naturales de unión a ADN de factores de transcripción homodiméricos son a veces asimétricos (Berg y von Hippel 1988; Sarai y Takeda 1989; Raumann *et al.* 1994; Pyles y Lee 1998; King *et al.* 1999; Winston *et al.* 2000;

Kalodimos *et al.* 2002; Ramirez-Carrozzi y Kerppola 2003) y pueden tener mayor afinidad que frente a sitios simétricos (Sarai y Takeda 1989; Winston *et al.* 2000). Además, sitios asimétricos de unión pueden curvarse en forma asimétrica al unirse a proteínas y/o conducir a cambios asimétricos en la estructura de la proteína de unión (Raumann *et al.* 1994; Pyles y Lee 1998; King *et al.* 1999; Ramirez-Carrozzi y Kerppola 2003), y revisado en (Brown 2006). Esto a su vez, puede regular la estructura del ADN lindante al sitio de unión y/o la unión de otras proteínas (Berg y von Hippel 1988; Pyles y Lee 1998; Ramirez-Carrozzi y Kerppola 2003). La asimetría en los sitios de unión de E2 se encuentra orientada de forma conservada y varía de sitio a sitio (**Figura II.2**). Se reportó que el desarrollo de la replicación es sensible a la orientación de sitios de unión de E2 modificada artificialmente (Newhouse y Silverstein 2001). En este trabajo proponemos que la orientación de sitios de unión de E2 naturales induce una torsión asimétrica del ADN y/o cambios en la estructura de E2C y juega un rol funcional de regulación de la unión de otras proteínas o dominios de E2 a sitios de ADN adyacentes.

Comparamos el estudio de sitios de alfapapilomavirus con los sitios de unión de E2 de betapapilomavirus, dada su creciente relevancia para el hombre. En efecto, el *genus* alfa compuesto por 72 tipos distintos es el más abundante de papilomavirus, aunque el *genus* beta con 42 tipos es el segundo más abundante de papilomavirus (Bernard *et al.* 2010) y 40 de ellos fueron aislados en humanos (Bernard *et al.* 2010). A diferencia de los alfapapilomavirus, todos los betapapilomavirus infectan tejido epitelial cutáneo y son causales de una gran variedad de enfermedades cutáneas, dentro de las cuales se destaca la enfermedad epidermodisplasia verruciforme, conocida como la enfermedad del *hombre árbol*, (Steger *et al.* 1988), y la fuerte asociación con el cáncer de piel inducido por exposición al sol (Forsslund *et al.* 2007). A medida que se analizan más muestras de tejidos cutáneos, se encuentran nuevos tipos pertenecientes al *genus* beta (de Villiers *et al.* 2009; Wood *et al.* 2010), revelando el interés existente en este *genus* en la comunidad científica de papilomavirus. Finalmente, a diferencia del *genus* alfa, aún no se han desarrollado en betapapilomavirus tratamientos preventivos como ser vacunas profilácticas. Esta breve descripción del contexto del *genus* cutáneo destaca el interés en discutir si las conclusiones obtenidas a partir del estudio de secuencia de unión de E2 en alfapapilomavirus pueden ser extrapolables al *genus* beta. Con este fin, en la **Figura 2** analizamos todos los sitios de ADN de unión a E2 y la hélice $\alpha 1$ de reconocimiento de ADN del *genus* beta.

En primer lugar, podemos observar que en betapapilomavirus, al igual que para el *genus* alfa, se mantiene conservada la secuencia de los dos hemisitios ACCG-CGGT (comparar **Figura II.3** con **Figura 2**). Sin embargo, la conservación de la asimetría presente en la secuencia espaciadora en los sitios del *genus* alfa (**Figura II.3**, posiciones 6-7-8 y 9) desaparece en el *genus* beta (**Figura 2**), insinuando que la torsión asimétrica del ADN es una característica de la regulación génica vía E2 en alfa-papilomavirus y no en betapapilomavirus. Por lo tanto, si la conservación del ‘código’ de lectura indirecta característico del *genus* alfa cambia en los sitios específicos de unión a E2 del *genus* beta, a diferencia de la conservación del ‘código’ de lectura directa que se mantiene igual, es de esperar que de cambiar la conservación de la secuencia proteica de reconocimiento en E2, resulten afectados aquellos residuos adaptados a la lectura indirecta de la interacción. Con el fin de discutir este aspecto, extrajimos las secuencias de la hélice $\alpha 1$ del alineamiento de secuencia de todas las proteínas E2 del *genus* beta para compararla con la correspondiente al *genus* alfa (**Figura II.1**), mediante el logo diferencial de dos muestras (**Figura 2**, abajo).

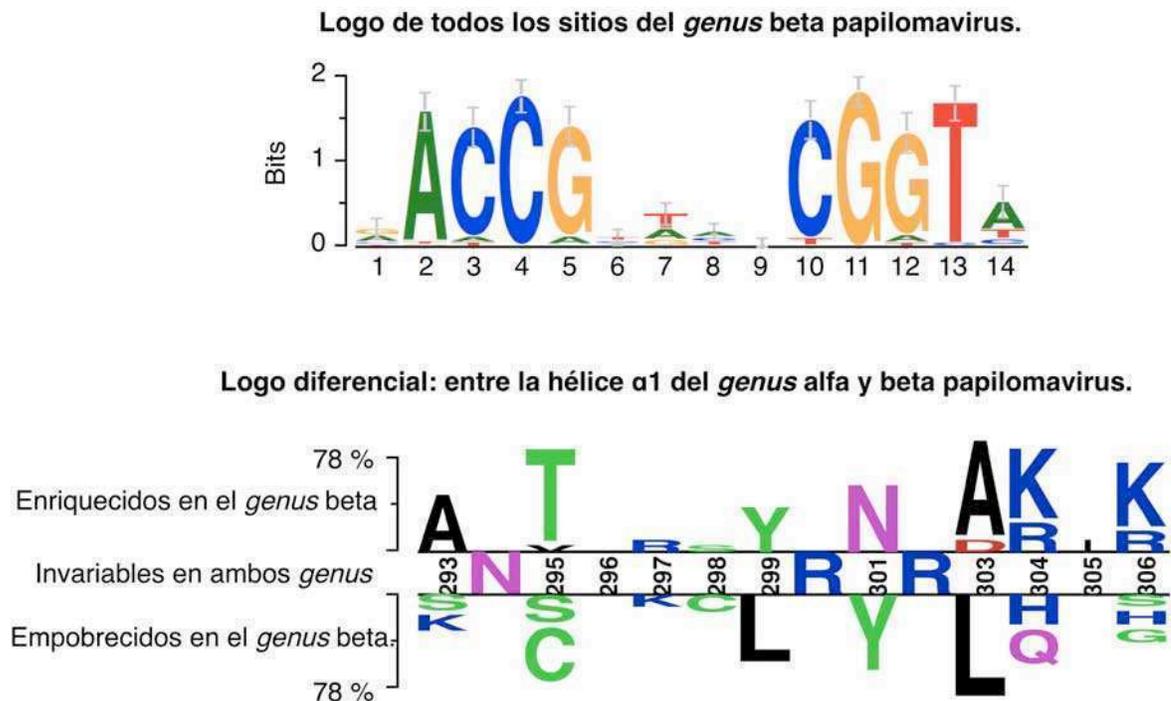


Figura 2. Logos de secuencia de la interacción E2-ADN en el *genus* beta papilomavirus. Arriba: Logo de secuencias (Crooks *et al.* 2004), del alineamiento de todos los sitios de ADN biológicos de unión de E2 del *genus* beta papilomavirus (comparar con **Figura II.3**). Abajo: Logo diferencial de dos muestras (Vacic *et al.* 2006), siendo el alineamiento de la hélice $\alpha 1$ de E2 del *genus* beta la muestra y el alineamiento de la hélice $\alpha 1$ de E2 del *genus* alfa el contraste (ver **Figura II.1**).

Al analizar en conjunto la conservación de la hélice $\alpha 1$ entre ambos *genus* (**Figura 2, abajo**), se evidencia que las diferencias en la conservación de secuencias involucradas en la interacción E2C-ADN involucra tanto el ADN como la proteína. La posición 301 presenta una tirosina conservada en la hélice $\alpha 1$ del *genus* alfa (**Figura II.1**), y una asparagina conservada en el *genus* beta (**Figura 2**). El $\Delta\Delta G$ entre la mutante Y301N y el dominio salvaje E2C-16 es de 2.04 kcal/mol (Ferreiro *et al.* 2005), indicando que la asparagina en la posición 301 debilita significativamente la interacción E2C-ADN en alfafapilomavirus. Suponemos que no ocurriría lo mismo en la interacción E2C-ADN del *genus* beta dado que la asparagina está conservada en esa posición (**Figura 2**). Nos preguntamos entonces si estas diferencias en la conservación de los residuos involucrados en la interacción E2-ADN, estarían compensando la falta de precurvatura en el ADN de los sitios de unión del *genus* beta. Se ha propuesto que las cargas positivas en la superficie electrostática de la región de reconocimiento de ADN de E2 promueven la curvatura del ADN (Hegde 2002). Un indicio de la hipótesis de compensación se asienta en la presencia de residuos de carga positiva conservados diferencialmente en betapapilomavirus, ver posiciones 304 y 306 (**Figura 2, abajo**). A futuro, estudios biofísicos realizados a partir de la proteína E2 de betapapilomavirus podrían responder estos interrogantes, que como hemos analizado, pueden aportar datos claves para la comprensión del mecanismo de lectura indirecta en este sistema y en general para interacciones proteína-ADN.

Los sitios BS1 y BS2 se encuentran codo a codo en el genoma de papilomavirus (**Figura II.2**). Estos dos sitios se encuentran flanqueados por una caja TATA y un sitio de unión de Sp1 a distancias muy conservadas (**Figura V.1**). Los dos sitios están involucrados en la regulación de la transcripción de las proteínas tempranas del virus (Thierry y Howley 1991; Dong *et al.* 1994; Demeret *et al.* 1997; Rapp *et al.* 1997; Steger y Corbach 1997; Stubenrauch *et al.* 1998; Newhouse y Silverstein 2001; Soeda *et al.* 2006), sugiriendo que actúan como una unidad funcional en alfa papilomavirus. Efectos cooperativos han sido observados para la unión de E2C-16 al DBS (Tan *et al.* 1994), los cuales pueden también involucrar a Sp1 y TBP. La inserción de una base entre BS1 y BS2 está relacionada con la oncogenicidad de ciertos tipos virales (**Tabla II.1**), sugiriendo que las distancias exactas dentro de esta unidad funcional puede tener una relevancia funcional. La distancia entre BS2 y BS3 también se encuentra restringida en 66 de los 72 tipos analizados (**Figura II.2**), sugiriendo que junto con BS1 extienden la unidad funcional involucrando además la replicación, junto con el sitio de unión de la proteína E1 (Chiang *et al.*

1992; Remm *et al.* 1992; Lu *et al.* 1993; Demeret *et al.* 1995; Sverdrup y Khan 1995; Stubenrauch *et al.* 1998; Newhouse y Silverstein 2001). La variabilidad de la posición de BS4 puede estar relacionada con la diversidad de sitios de unión de proteínas celulares en la región flanqueada por BS4 y BS3 (Garcia-Vallve *et al.* 2006). Finalmente, en la **Figura 3**, observamos que el patrón de distancias restringidas entre los cuatro sitios Sp1, BS1, BS2 y caja TATA es una característica propia del *genus* alfa y no de betapapilomavirus. Por ejemplo, en el *genus* beta la distancia entre el BS1 y el BS2 alcanza hasta 24 bases, evidenciando que el bloque funcional del *genus* beta involucra diferentes vías de regulación genómica.

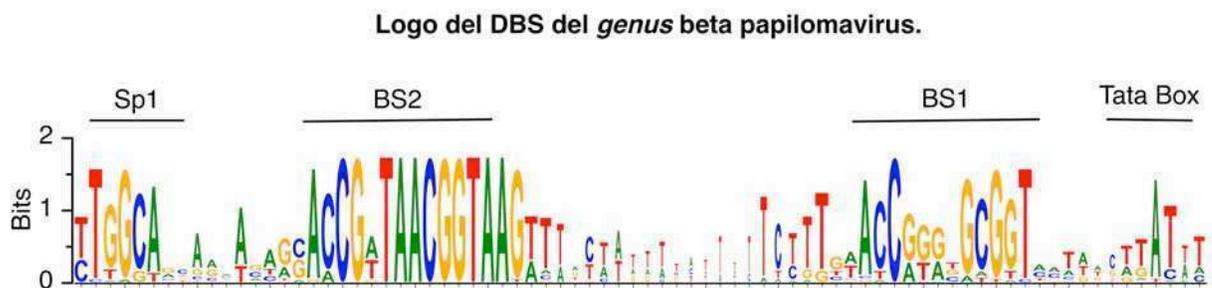


Figura 3. Logos de secuencia del DBS en el *genus* beta papilomavirus. Logo de secuencias (Crooks *et al.* 2004), de todos los elementos de respuesta *cis* del promotor de E6 en el *genus* beta papilomavirus (comparar con **Figura V.1**).

Los dinucleótidos CG se encuentran sub-representados en el genoma de HPV alfa (Badal *et al.* 2003), sugiriendo que el virus ha eliminado sitios de metilación perjudiciales. Sin embargo, hay una presión evolutiva a favor de por lo menos un sitio de metilación en los sitios de unión de E2 (**Figura II.3**). Esta presión varía entre los cuatro sitios, sugiriendo que podría estar relacionada con la función de cada sitio. Curiosamente, la región reguladora de alfa papilomavirus posee sitios de unión para proteínas del huésped involucradas en procesos de desmetilación cercana a uno de los sitios de unión de E2 (Gloss *et al.* 1987; Tan *et al.* 1994; Kirillov *et al.* 1996; Matsuo *et al.* 1998; Fontaine *et al.* 2000; Thomassin *et al.* 2001), como el receptor de glucocorticoides, Sp1 y NF- κ B. Además, la proteína E2 puede actuar como un cofactor del receptor de glucocorticoides (Wu *et al.* 2007). A partir de estos datos podemos deducir que los alfa papilomavirus han integrado la metilación en los sitios de unión de E2 influidos por regulaciones mediadas por la célula huésped a lo largo del ciclo infeccioso del virus, convirtiendo este usual mecanismo de defensa del huésped en una regulación extra del virus. Esta regulación puede desarrollarse simplemente mediante cambios en la jerarquía de

afinidades al metilarse los sitios de unión de E2 (Thain *et al.* 1996). Notablemente, estos cambios pueden ser diferentes en tipos de alto-riesgo y tipos cutáneos debido a la falta de sitios de metilación (**Figura II.3** y **Tabla II.1**).

Las afinidades de los cuatro sitios de unión de ADN de la proteína E2 en los 72 tipos del *genus* alfa puede en principio constituir un continuo de muchas jerarquías diferentes. Sin embargo, las energías predichas de la unión pueden ser agrupadas en solo seis jerarquías prototípicas, bien diferenciadas una de otra (**Figura II.5**), de las cuales, las tres jerarquías más comunes incluyen el 85% de los tipos. Esta simplicidad inesperada implica que sólo un pequeño número de jerarquías bien definidas hacen funcional al virus. Este hecho genera dos desafíos para nuestra comprensión de los alfapapilomavirus. Primero, cómo los roles reguladores de la transcripción y replicación a lo largo del ciclo infeccioso del virus, diagramados finamente por los cuatro sitios de unión de E2, pueden estar determinados por las jerarquías de afinidades. Segundo, cuál sería el mecanismo por el cual estas jerarquías se relacionan con el desarrollo del cáncer.

El *genus* alfa incluye no sólo a todos los tipos de HPV genitales, sino que también incluye doce tipos cutáneos (Bernard 2006; Bravo y Alonso 2007). Cerca de un tercio de los tipos cutáneos de *genus* beta poseen un quinto sitio para la unión de E2C en zonas lindantes con el sitio BS4 (Ensser *et al.* 1990), los cuales fueron tenidos en cuenta al construir el logo de todos los sitios de unión de E2 del *genus* beta (**Figura 2**). La presencia de un quinto sitio coincide con los sitios adicionales que se descubrieron en seis tipos de alfapapilomavirus (Sanchez *et al.* 2008). Sin embargo los tipos con el sitio adicional no se relacionan filogenéticamente entre sí (Bernard 2006; Bravo y Alonso 2007) y ninguno de ellos causa verrugas cutáneas. Una fina reexaminación de los sitios de unión de E2 en el *genus* beta papilomavirus revela que su función (Stubenrauch *et al.* 1996), posición relativa (Guido *et al.* 1992) y secuencias consenso son significativamente diferentes con ambos, cutáneos y mucosos de alfapapilomavirus (**Figuras 2** y **3**). Especulamos que los alfapapilomavirus pueden haber desarrollado su propia estrategia para infectar tejido epitelial no mucoso. Ambos, la unión de factores de transcripción celulares (Garcia-Vallve *et al.* 2006) y las características de los sitios de unión de E2 asociados con tipos cutáneos de alfapapilomavirus pueden formar parte de esta estrategia.

La proteína E2 es conocida como regulador maestro de la replicación de papilomavirus, transcripción y segregación. La identificación de propiedades de E2 asociadas con tipos de alto

riesgo destaca su importancia como supresor tumoral. Nuestros resultados confirman las hipótesis previas sobre la importancia de la jerarquía de afinidades de sitios de unión de E2 (Alexander y Phelps 1996; Dell *et al.* 2003) y de los sitios de metilación (Badal *et al.* 2003; Kim *et al.* 2003; Badal *et al.* 2004; Kalantari *et al.* 2004; Wiley *et al.* 2005; Bhattacharjee y Sengupta 2006), siendo además, extrapolables a todos los alfapapilomavirus. Podemos ahora sumar la distancia entre los sitios BS1 y BS2 a la lista de genotipos asociados con el desarrollo de la enfermedad (Hou *et al.* 2002; Ferreira y Prat Gay 2003; Hooley *et al.* 2006; Parish *et al.* 2006; Klucsevsek *et al.* 2007). Resulta interesante analizar la interacción entre estas propiedades de E2 y las propiedades moleculares de las proteínas oncogénicas E6 y E7, conocidas por estar asociadas con cáncer. El conocimiento generado en este trabajo podría ayudar además a entender el comportamiento epidemiológico de las variantes moleculares de HPV (Sichero y Villa 2006).

Estudios termodinámicos, estabilidad y reconocimiento de ADN.

La comprensión de los mecanismos moleculares de las interacciones proteína-ADN es un desafío fundamental cuyo abordaje deriva en innumerables aplicaciones prácticas. Para este fin, el estudio de proteínas homólogas es una herramienta clave. Principalmente porque las proteínas homólogas comparten información valiosa incluida en la conservación de sus secuencias. Homología es sinónimo de origen en común, vale la aclaración dado que en la literatura se encuentran numerosos casos que utilizan el concepto en forma confusa (Reeck *et al.* 1987). Al no ser un término cuantitativo no es correcto calificar como “altamente homólogas”, “significativamente homólogas” o “x % de homología” a secuencias, genes o proteínas, sin embargo se pueden encontrar en la literatura 300 ejemplos con semejantes calificaciones, aún en el título de artículos (Nakamura *et al.* 2010; Peng *et al.* 2010)). Cerrando el paréntesis, en nuestro caso, las proteínas homólogas E2 de papilomavirus comparten los pilares fundacionales que garantizan el correcto desarrollo del ciclo infeccioso del virus. Comúnmente los estudios entre proteínas de una misma familia se focalizan en comprender el mecanismo de plegamiento (Gunasekaran *et al.* 2001; Nickson *et al.* 2010), o el mecanismo funcional (por ejemplo (Robbins *et al.* 2004)), siendo escasos los trabajos que los traten en simultáneo y en las mismas condiciones. Logramos obtener un grupo de cinco dominios homólogos E2C con alto grado de pureza, los cuales presentan una variabilidad de secuencia suficiente como para extrapolar el grupo al resto de los dominios de unión a ADN de alfapapilomavirus (**Figura IV.1**). En la

discusión del Capítulo anterior quedó claro que extrapolar el grupo al *genus* beta sería imprudente (**Figura 2A**). Estudiamos los aspectos bioquímicos básicos del grupo, es decir características generales de su estructura primaria, secundaria, terciaria y cuaternaria (**Figura IV.2**), los cuales permitieron respaldar los modelos plegamiento empleados para interpretar los datos de desnaturalización al equilibrio (**Figura IV.3**) y los datos cinéticos de desnaturalización (**Figura IV.4**). Estas a su vez, garantizaron el estado dimérico de todos los dominios del grupo otorgando confianza a los resultados de los ensayos funcionales conformacionales (**Figura IV.5**) y termodinámicos (**Figura IV.6**). Gracias a la caracterización funcional previa utilizando el dominio E2C-16 (**Figura III.1**) y al conocimiento acumulado en nuestro laboratorio, (Ferreiro *et al.* 2000; Ferreiro 2003; Dellarole 2005; Ferreiro *et al.* 2005; Cicero *et al.* 2006; Dellarole *et al.* 2007; Ferreiro *et al.* 2008; Eliseo *et al.* 2009; Sanchez *et al.* 2010), logramos realizar todos los experimentos en un mismo tampón químico, el fosfato de sodio. Por lo tanto, los resultados expuestos en el Capítulo III pueden ser analizados en conjunto y en forma comparativa entre los cinco dominios.

A pesar de la poca diferencia que existe en la estructura tridimensional de los dominios E2C (E2C-6: PDB libre: 1r8h y PDB complejo: 2ayg, (Dell *et al.* 2003; Hooley *et al.* 2006); E2C-16: PDB: 1by9 y 1r8p, (Hegde y Androphy 1998; Nadra *et al.* 2004); E2C-18: PDB libre: 1f9f y PDB complejo: 1jj4, (Kim *et al.* 2000) y E2C-31: PDB: 1a7g, (Bussiere *et al.* 1998)) observamos grandes diferencias en la señal de CD en el UV lejano reportando la estructura secundaria en condiciones nativas (**Figura IV.2**). Sin embargo, se ha reportado que cambios mínimos en la secuencia generan grandes cambios en el espectro de CD en el UV lejano (Ferreiro 2003; Ferreiro *et al.* 2005) indicando la sensibilidad de la señal de CD en la estructura secundaria del dímero E2C; y dada la gran variabilidad de secuencia del grupo de las cinco E2C es esperable encontrar variaciones, aunque no tantas, indicando una particularidad más propia de este dominio dimérico de barril β . En efecto, no encontramos diferencias en las señales que reportan la estructura terciaria medida por CD en el UV cercano y en estructura cuaternaria medida por DLS (**Figura IV.2**), más allá de la diferencia aportada por la presencia diferencial de residuos aromáticos (**Figura IV.1**), en línea con las evidencias estructurales.

Así como anteriormente se había observado la presencia de estructura residual en el estado desnaturalizado de E2C-16 (Mok *et al.* 2000), curiosamente en el presente estudio observamos lo mismo para los cinco dominios. En efecto, en todos los casos podemos notar

pendientes en los estados desnaturalizados monitoreados por ambas señales fluorescencia y CD (**Figura IV.3**). Dicha estructura remanente puede entonces explicar el origen de las diferencias observadas en valor del parámetro m_{eq} , el cual se correlaciona empíricamente el cambio en la superficie expuesta al solvente (Myers *et al.* 1995). Entonces podemos inferir que dicha estructura residual presente en el estado desnaturalizado es distinta entre las cinco proteínas homólogas, pero dado el alto grado de divergencia es difícil determinar que región en particular está involucrada. Los estudios cinéticos mostraron que el estado de transición de la reacción de desnaturalización presenta mayoritariamente estructura nativa en forma muy similar entre los cinco dominios (**Figura IV.4**), por lo tanto, sorpresivamente confirmamos experimentalmente que las diferencias se encuentran en el estado desnaturalizado. En conjunto, queda claro que los cinco dominios se comportan en condiciones de desnaturalización de acuerdo a un modelo simple de dos estados, y que las proteínas homólogas son lo suficientemente estables como para garantizar su naturaleza dimerica en los ensayos de discriminación de secuencias de ADN medidos por ITC.

Al igual que en el Capítulo III para el dominio E2C-16 (**Figuras III.4 y III.5**), caracterizamos la interacción de la unión de los cinco dominios E2C a tres sitios de ADN: BS1, BS2 y BSR (**Figuras IV.5 y IV.6**). ¿Por qué elegimos estos tres sitios de ADN? Estudios de migración en gel y en solución indicaron previamente que las secuencias de *tramo-A*, compuestas por adeninas y timidinas, se curvan hacia el surco menor del ADN, es decir que son secuencias de ADN intrínsecamente curvadas (Crothers *et al.* 1990). Esta propiedad de precurvatura fue comprobada cristalográficamente por un lado por la presencia de precurvatura en sitios de unión a E2 de espaciador del tipo *tramo-A* (Rozenberg *et al.* 1998; Hizver *et al.* 2001), y por otro lado, por la ausencia de precurvatura en sitios de unión a E2 diseñados sin el *tramo-A* (Rozenberg *et al.* 1998). Dichas observaciones cristalográficas fueron luego confirmadas en solución mediante experimentos en geles de poliacrilamida nativos (Zimmerman y Maher 2003). El ADN con *tramo-A* en su secuencia espaciadora presenta un ángulo de curvatura de diez veces mayor en grados que la observada para el ADN sin el *tramo-A* BSR (**Figura 4 Introducción**) (Zimmerman y Maher 2003). Algunos autores sugieren que esta precurvatura genera ventajas entrópicas para la unión de E2C (Kim *et al.* 2000; Dell *et al.* 2003). El evento de precurvatura es llamado en general lectura indirecta de la interacción proteína-ADN (Shatzky-Schwartz *et al.* 1997; Crothers 1998; Hizver *et al.* 2001; MacDonald *et al.* 2001;

Barbic *et al.* 2003; Xi *et al.* 2010). Como estudiamos en detalle en el Capítulo II, la secuencia espaciadora de los sitios naturales BS1 y BS2 es del tipo *tramo-A*, mientras que el sitio artificial BSR carece de *tramo-A* y corresponde a la misma secuencia que se utilizó en los trabajos recién mencionados (Rozenberg *et al.* 1998; Zimmerman y Maher 2003; Hooley *et al.* 2006).

De acuerdo a la evidencia cristalográfica, la curvatura adicional del ADN generada por la unión de E2C es de unos 10 a 20 grados más que la precurvatura del ADN libre (Kim *et al.* 2000; Hooley *et al.* 2006). Existen también evidencias de la curvatura adicional a la precurvatura medidas en solución por CD (Lima y Prat Gay 1997; Ferreiro *et al.* 2000). Estudiamos la capacidad diferencial de los cinco dominios homólogos de inducir cambios conformacionales en el sitio de ADN BS2 (**Figura IV.5**). La señal de CD como reportera de estructura de ADN (Baase y Johnson 1979; Gray *et al.* 1995; Gray 1996), mostró que la forma B del ADN se conserva como tal al unir los dominios, coincidiendo con los datos estructurales previamente medidos para los dominios E2C-6 y E2C-18 (Rozenberg *et al.* 1998; Kim *et al.* 2000; Hizver *et al.* 2001; Hooley *et al.* 2006). Los cambios estructurales en el BS2 monitoreados de acuerdo a la magnitud de la señal de absorbancia y de CD presentaron ligeras variaciones entre las cinco proteínas (**Figura IV.5**). Se ha sugerido que la habilidad de E2C en promover la curvatura del ADN se correlaciona con las cargas positivas de la superficie electrostática de la proteína (Hegde 2002). La proteína E2C-18 presenta mayor carga positiva que la E2C-16, coincidente con una mayor curvatura del ADN (Hooley *et al.* 2006). Una mirada fina de los espectros medidos indica que la magnitud de la señal de CD y absorbancia correspondientes al dominios E2C-18 son mayores que el resto (**Figura IV.5**), con mayor claridad para BS1 (**Figura V.7**). Sin embargo, consideramos que globalmente los cambios en magnitud de señal no son significativos en nuestras condiciones experimentales, en línea con un *pI* similar entre los cinco dominios (**Tabla IV.1**). Es de esperar, finalmente, que la inducción del cambio estructural del ADN se asiente en residuos altamente conservados entre las cinco homólogas y probablemente conservados en todo el *genus* alfa.

Las diferencias en la disección termodinámica de la interacción entre los cinco dominios y los tres sitios de ADN (**Figura IV.6**), va a estar relacionada principalmente con la capacidad de los dominios en adaptarse a la falta de precurvatura del ADN. Antes de discutir este aspecto, resulta crucial investigar si el modo de unión a los distintos sitios de ADN se encuentra conservado para las cinco proteínas homólogas. Para ello estudiamos la compensación entrópico-

entálpica como describimos en el Capítulo III únicamente para la proteína E2C-16 y los siete sitios distintos de unión a ADN ensayados (**Figura III.6**). Ya que buscamos analizar el modo de unión en la interacción con ADN de 5 dominios proteicos que representan todo un *genus*, decidimos aumentar el número de datos a analizar, con el fin de obtener una conclusión abarcadora del sistema E2C-ADN. En la **Figura 4** graficamos los parámetros de ΔH en función de $-\Delta S$ y de ΔH en función de ΔG correspondientes a los resultados presentados en las **Tablas III.1, III.2 y IV.4**, en (Dellarole *et al.* 2007), en (Cicero *et al.* 2006) y datos no publicados (ver leyenda, **Figura 4**). De esta forma recopilamos 30 interacciones involucrando al complejo E2C-ADN y variando la molecularidad de la proteína (dos versiones monomerizadas de E2C-16), la secuencia de la proteína (cinco homólogas con un rango en identidad de secuencia del 45 % al 77% y una doble mutante en la hélice $\alpha 1$), la secuencia del ADN (variando hasta 9 de las 14 bases del sitio de unión) y la composición del tampón químico de medición (con una diferencia en la fuerza iónica de 0.6M) respetando la temperatura de medición en 25°C y el pH en ~ 7 .

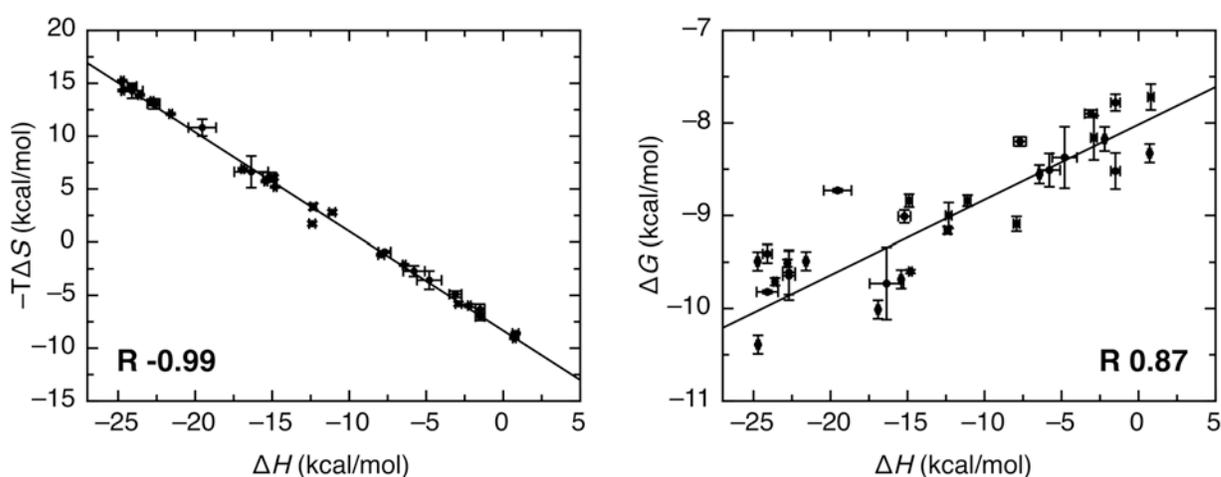


Figura 4. Conservación del modo de unión a ADN por E2C. Compensación entrópico-entálpica y correlación entre la entropía y la energía libre de Gibbs para el complejo E2C-ADN realizados en 30 experimentos de ITC variando tanto la fuerza iónica y la naturaleza química de la solución amortiguadora como la secuencias de ambas especies del complejo. En el tampón químico fosfato pH 7 y a 25°C: 5 complejos E2C-16:BS2 variando la concentración de fosfato (**Tabla II.1**), 6 complejos E2C-16-ADN variando la secuencia de ADN a los sitios BS1, BS3, BS3*, BS4, BSR y Bov (**Tabla II.2**), 12 complejos E2C-ADN variando la secuencia de ADN a BS1, BS2 y BSR y las proteínas E2C-11, E2C-18, E2C-31 y E2C-45 (**Tabla III.4**), 2 complejos E2C-17:BS2 con E2C monomerizada mediante un conector de 6 y 12 residuos (Dellarole *et al.* 2007), 2 complejos E2C-16:BS2 variando el largo del oligonucleótido a 18 y 24 bases y a pH 5.6 (Cicero *et al.* 2006), 2 complejos E2C-16:BS2 en el tampón químico Bis tris pH 7 200 mM NaCl con E2C-16 salvaje y doble mutante Y301A-K297A (datos no publicados) y 1 complejo E2C-16:BS2 en el tampón químico MES pH 6 200 mM NaCl (datos no publicados). Las líneas corresponden a la regresión lineal de los datos, panel izquierdo: pendiente = -0.93 ± 0.1 y ordenada al origen = -8.3 ± 0.1 , panel derecho pendiente = -0.081 ± 0.002 y ordenada al origen = -8.1 ± 0.1 cuyo coeficiente de correlación se indica en cada panel.

Sorprendentemente, al igual que lo visto únicamente para el dominio E2C-16 (**Figura III.6**), a pesar de una extensísima variabilidad de condiciones, secuencias, características energéticas y conformacionales, existe una compensación entrópico-entálpica isotérmica indicando que el modo de unión es el mismo en todos los complejos analizados en la **Figura 4**. La correlación lineal entre el ΔH y el ΔG ($R=0.87$), dos parámetros que se obtienen en forma independiente en los ensayos de ITC, y la calidad de los datos, con errores provenientes del desvío estándar de al menos dos experimentos de variable concentración proteica y nucleotídica, garantizan que las correlaciones reflejan un fenómeno físico y no un artefacto estadístico (Leffler y Grunwald 1963; Jen-Jacobson *et al.* 2000). En contraste con este comportamiento en E2C, con sólo un cambio de base en el sitio de unión específico de la enzima de restricción BstY1 el modo de unión cambia drásticamente (Townson *et al.* 2007). Alternativamente, como estudiamos en la dependencia de la energía de unión con el pH (**Figura III.7**), la presencia de eventos de protonación converge en una correlación elíptica en los parámetros termodinámicos de la unión (**Figura III.8**). Proponemos que el elevado grado de robustez del complejo E2C-ADN puede estar relacionado con un requisito fisiológico para que el dominio E2C forme complejos con cuatro sitios distintos (Prat Gay *et al.* 2008).

En este panorama de adaptabilidad del dominio E2C frente al reconocimiento de ADN, apreciamos que los cinco dominios poseen la misma capacidad de discriminación de secuencias (**Figura IV.7**). La afinidad al ADN específico, BS1 y BS2, resultó ser fuertemente mayor que la afinidad al sitio BSR ($\sim >1$ kcal/mol, (**Figura IV.8**), en excelente acuerdo con su carácter de sitio recto. Recordamos que las dos hélices $\alpha 1$ de la proteína E2C unen al ADN encastrándose en los surcos mayores del ADN, de secuencia conservada y pseudopalindrómica ACCG y CGGT (**Figura II.1**). Entre ambos surcos mayores se encuentra el surco menor del ADN, cuya secuencia espaciadora suele ser rica en la base A y/o T. Los dos surcos mayores se aproximan entre sí gracias a la curvatura del surco menor (**Figura II.1**). El sitio BSR posee un surco menor que se encuentra extendido, sin curvatura, dada la composición pobre en bases A y/o T. En efecto, la extensión del surco menor en el sitio de ADN BSR alcanza a medir el doble que la distancia del surco menor de los sitios con el espaciador de *tramo-A* como ocurre en el BS1 y en el BS2 (Hooley *et al.* 2006). Llamativamente, la contribución energética que gobierna la interacción en el BSR es la entropía, a diferencia de los otros dos sitios donde gobierna la entalpía (**Figura IV.7**). Al anularse la lectura indirecta en el BSR, es decir la curvatura del ADN,

podríamos asignar a grandes rasgos que la diferencia de energía entálpica observada para la unión al BSR se origina principalmente en la lectura directa, la cual sería entrópicamente favorable (**Figura IV.7**). Por lo tanto, y suponiendo que la falta de curvatura en el complejo no modificaría el número de contactos directos entre la proteína y el ADN, en la unión a ADN específico que exhibe una curvatura significativa en el ADN, la diferencia de energía entálpica se originaría principalmente en la desolvatación de la interfaz, producto de la torsión del ADN hacia el surco menor. Resultaría interesante realizar a futuro experimentos que midan la dependencia de la energía de unión con la fuerza iónica en la formación del complejo E2C-BSR, y compararlo con el complejo E2C-BS2 (**Figura III.3**). Es entonces, gracias a que medimos la interacción entre un sitio de ADN que carece de precurvatura y cinco proteínas homologas, que podemos especular sobre el origen de las contribuciones entálpicas en la formación del complejo entre E2C-ADN. La comprensión inteligible de la disección termodinámica de interacciones entre proteínas y ADN puede ser de potencial utilidad para el diseño de proteínas que reconozcan una dada secuencia de ADN en forma específica con fines terapéuticos.

Bases termodinámicas para el reconocimiento cooperativo de ADN.

De manera de complementar la estructura y la biología de orígenes de replicación de ADN eucarióticos, y en particular aquellos de virus de ADN doble cadena, realizamos una disección termodinámica y conformacional del complejo entre el dominio de reconocimiento de ADN y regulador maestro, E2 y dos sitios de ADN diana adyacentes (**Figura V.1**). El análisis de los datos para la formación del complejo ternario se estudia usualmente mediante técnicas de EMSA (Tan *et al.* 1994) y de ensayos espectroscópicos (Weinberg *et al.* 2004; Dragan *et al.* 2008). Sin embargo, cuando el modelo de unión es complejo, como el presente trabajo, donde BS1 y BS2 no son equivalentes (diferente composición de secuencia) ni independientes (diferente afinidad individual), se debe tratar de utilizar el modelo más simple posible. Para estudiar la interacción cooperativa de E2C y DBS empezamos utilizando EMSA y espectroscopía de fluorescencia, aunque dada la intrincada naturaleza de nuestro modelo de unión (comparar **Figura III.2** con **Figura V.3**), no pudimos interpretar la población de especies formadas más allá de la verificación de la estequiometría (**Figura V.2**). Esto implica que la descripción cuantitativa de nuestros complejos ternarios no es posible basándonos en la señal de fluorescencia, por lo tanto nos concentramos en ITC.

El hecho de haber medido las entalpías de los sitios individuales en forma independiente y adecuada (**Figuras III.3 y IV.4**), nos permitió abordar el análisis de unión del complejo ternario E2C:DBS:E2C por ITC. Logramos ajustar globalmente las titulaciones directas y reversas del sitio en tándem (**Figura V.4**) (Houtman *et al.* 2007). Es llamativo que los valores entálpicos adicionales (**Figura V.6**), debidos a la unión de una segunda molécula de E2C estando la primera ya unida al DBS, son muy pequeños en forma relativa a la unión de los sitios aislados (**Tablas III.2 y IV.4**), acercándose al ruido instrumental. Sin embargo, dichos valores de entalpía adicional para la unión del DBS fueron los esperados de acuerdo al análisis cualitativo de las curvas (**Figura V.3**), donde apreciamos que la suma o el promedio de las entalpías de los sitios aislados coincidía con las titulaciones directas o reversas medidas para el sitio doble. A pesar de que la escala de los parámetros termodinámicos de cooperatividad obtenidos es pequeña, consideramos firmemente que son sólidos. Primero, la entalpía adicional para la unión de E2C a dos sitios en tándem coincide para cada dominio ensayado en el ajuste individual y en el global, así como los obtenidos en mediciones adicionales (**Tabla V.1**). Segundo, los ajustes fueron restringidos a los parámetros de sitios aislados provenientes cada uno de varias medidas individuales ajustadas en forma global (ver Capítulos III y IV). Y tercero, el valor de K_{coop} del mismo sistema pero proveniente de otra técnica y grupo de trabajo (Tan *et al.* 1994) coincide con los obtenidos por en la presente tesis, reforzando nuestro abordaje y conclusiones. Vale esta aclaración dada la intrincada naturaleza de la interpretación de la cooperatividad por ITC ((Bains y Freire 1991; Freire *et al.* 2009), revisado por (Brown 2009)) y que en la literatura las mediciones e interpretación vía ITC de sistemas complejos, como es el caso de complejos ternarios no-idénticos y heterogéneos, suelen ser incompletos, es decir sin ajustes globales entre titulaciones directas y reversas ni el uso de modelos adecuados (por ejemplo: (Schumacher *et al.* 2002; Lacal *et al.* 2008)), imposibilitándonos comparar nuestros datos con otros sistemas.

La formación de complejos binarios de BS1 y BS2 con E2C-16 se encuentra asociada con cambios pequeños pero significativos en la señal de CD a 225 nm, de aproximadamente 1 mdeg y cambios de 0.04 y 0.06 en unidades de absorbancia a 270 nm (**Figura V.7**), de acuerdo con resultados medidos en solución anteriormente (Lima y Prat Gay 1997; Ferreiro *et al.* 2000). Los cambios espectrales al unirse el dominio E2C al sitio DBS difiere de aquellos esperados para el caso de una unión independiente, con unos 4 mdeg a 275 nm para el CD y unos 0.3 unidades de absorbancia a 270 nm, lo cual indica un cambio sustancial (**Figura V.8**). Estos resultados

sugieren que los cambios estructurales adicionales en el DBS asociados con la unión de E2C-16 son mayores que aquellos observados al unirse a los sitios aislados, de forma tal que la estructura del complejo ternario difiere considerablemente del modelo representado en la **Figura V.1**. Cabe señalar que (1) la forma del espectro de dicroísmo circular muestra que la distorsión del ADN de forma-B es pequeña, (2) el incremento en la señal de CD a 275 nm sugiere una disminución del ángulo de torsión de la hélice de ADN y un incremento en ángulo de giro de los pares de base (Johnson *et al.* 1981), y (3) que el incremento en señal de absorbancia es indicativo de desapareamiento de bases, del inglés *unstacking*, (Gray 1996; Di Pietro *et al.* 2003). Desde esta evidencia, sugerimos que el cambio en la estructura del ADN involucra perturbaciones significantes de la conformación de las bases nucleotídicas, pero pequeñas variaciones en la estructura de la cadena principal, dado que el CD indica que la forma-B se mantiene intacta. Visualizamos estos cambios en el ADN DBS como la formación de pequeñas burbujas que no perjudican la conformación global de la cadena principal del ADN, pero este evento primario es crucial para la iniciación de la actividad de la helicasa y el subsiguiente ensamblado de la maquinaria celular de replicación de ADN. Nuestros experimentos cinéticos muestran que existe una pequeña barrera de energía para que ocurra este cambio en la estructura, dado que los cambios ocurren en tiempos muy cortos, escala del submilisegundo. Desde el punto de vista funcional, la unión simultánea de dos moléculas de E2C-16 al DBS y los importantes cambios conformacionales generados, podrían asistir al veloz desenrollamiento de la doble hélice al iniciarse el progreso de la orquilla de replicación (Chiang *et al.* 1992; Lu *et al.* 1993; Sverdrup y Khan 1995; Alexandrov *et al.* 2010), podrían ayudar a reprimir la transcripción mediante el desplazamiento de SP1 y TBP u a otras proteínas reguladoras (Thierry y Howley 1991; Tan *et al.* 1992; Dong *et al.* 1994; Alexandrov *et al.* 2010) y/o podrían facilitar interacciones de largo alcance vía formación de rulos de ADN a través de su dominio N-terminal (Hernandez-Ramon *et al.* 2008; Sim *et al.* 2008).

La cooperatividad de unión descrita por la K_{coop} es similar para E2C-11 y E2C-16. Sin embargo, la cooperatividad es entálpicamente dirigida para E2C-11 y entrópicamente dirigida para E2C-16 (**Figura V.6**). Además, los cambios espectroscópicos asociados son claramente diferentes entre los dos dominios (**Figura V.8**). Podemos entonces concluir que de acuerdo a nuestra apreciación, describir las reacciones de unión de E2C-11 y E2C-16 al DBS como ‘similarmente cooperativas’, no sería una descripción adecuada o completa. En cuando a la K_{coop}

de E2C-18, la unión al DBS es aditiva, no cooperativa. Evaluando únicamente el parámetro K_{coop} , se tendería a describir la reacción de E2C-18 como si la unión de la primera molécula resultaría invisible para la unión de la segunda molécula al DBS (**Figura V.6**). Sin embargo, la falta de cooperatividad para la E2C-18 se debe a cambios en entalpía y entropía que se compensan, y a una asociación considerable de cambios en la estructura del ADN que va más allá de lo esperado para una unión aditiva (**Figura V.8**). Entonces, concluimos que describir la unión de E2C-18 a DBS simplemente como no-cooperativa no resulta en una apreciación completa: mientras que los cambios en ΔG son menores al 5%, las diferencias en los componentes entálpicos y entrópicos son mucho mayores y hasta de signo opuesto.

Por otra parte, la interacción E2C-ADN está determinada a grandes rasgos por las cadenas laterales de seis residuos altamente conservados en la hélice $\alpha 1$, involucrados en la lectura directa de las bases de ADN, **Figura II.1** (Sanchez *et al.* 2008), las cuales entre los tres dominios estudiados son idénticas (**Figura IV.1**). De acuerdo con esta particularidad, las diferencias de la disección termodinámica de la unión de las tres proteínas, E2C-11, 16 y 18, a los sitios que componen el DBS, BS1-2 (**Figura IV.8**), son muy similares, difiriendo en energía de unión en ~ 0.5 kcal/mol (**Tabla IV.4**). Este resultado contrasta con la variable K_{coop} para la asociación al DBS, mostrando una unión sinérgica para E2C-11 y E2C-16 pero de contribuciones energéticas opuestas y una unión aditiva para E2C-18 de contribución energética semejante a la de E2C-16 (**Figura V.6** y **Tabla V.1**). Entonces, la región de E2C involucrada en la formación del complejo ternario con el DBS, debe localizarse en una región de secuencia variable entre los tres dominios y expuesta al solvente.

Experimentos preliminares al momento de finalización de la escritura de la presente tesis, y por ello no incluidos en resultados, muestran diferencias entre los dominios E2C-11, 16 y 18, que podrían explicar este comportamiento termodinámico diferencial para el reconocimiento cooperativo. En la **Figura 5**, presentamos el perfil de elusión de *Superdex 75* de E2C-11 (en azul), E2C-16 (en rojo) y E2C-18 (en verde). El tamiz molecular es el último paso en la purificación de los dominios, el cual se realiza a temperatura ambiente, a elevadas concentraciones proteicas ($> 100 \mu\text{M}$) y en un tampón determinado (ver materiales y métodos). Para todos los dominios purificados, como ejemplificamos para los dominios E2C-16 y E2C-18 (**Figura 5A**), el peso molecular observado de acuerdo al perfil de elusión del tamiz molecular se corresponde con el esperado para un estado dimérico. Esta característica fue además confirmada

por DLS (**Figura IV.2**), y por los estudios de plegamiento (**Figuras IV.3 y IV.4**). Sin embargo y sorprendentemente, el peso molecular observado de acuerdo al perfil de elusión del tamiz molecular para E2C-11 presenta un pico adicional correspondiente al peso molecular de dos dímeros, o tetrámero (**Figura 5A**).

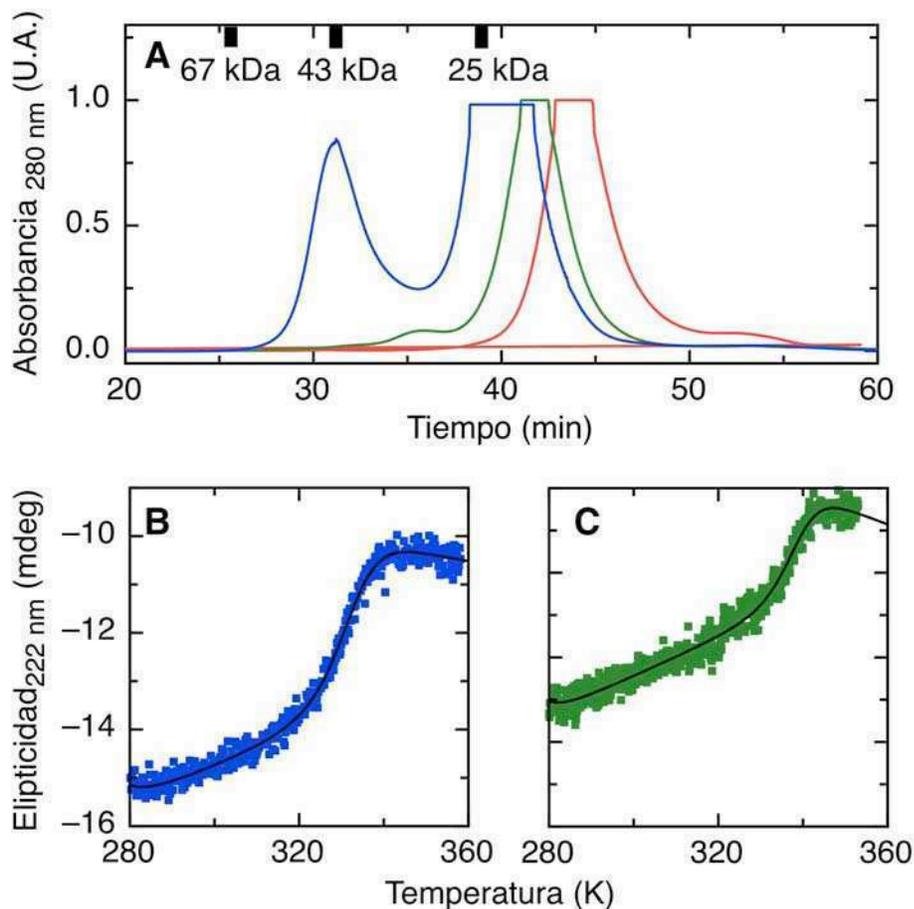


Figura 5. Particularidades observadas entre los homólogos E2C-11, E2C-16 y E2C-18. (A) Perfil de elusión para E2C-11 (azul), E2C-16 (rojo) y E2C-18 (verde) por tamiz molecular (*Superdex 75*), a una concentración proteica $>100 \mu\text{M}$, las barras negras se corresponden con el peso molecular de proteínas estándar. (B) Desnaturalización térmica de E2C-11 y (C) E2C-18 medida por señal de CD a 222 nm a una velocidad de calentamiento de $4^\circ\text{C}/\text{min}$, a una concentración proteica de $10 \mu\text{M}$. Línea negra, ajuste del modelo de disociación de dos estados a los datos: $T_{m\text{E2C-11}} = 58^\circ \pm 0.3\text{C}$ y el $T_{m\text{E2C-18}} = 65^\circ \pm 0.5\text{C}$. En ambos casos la desnaturalización es reversible. Tampón químico: 50mM NaCl, 10mM AcNa pH 5.6, 1 mM DTT.

El análisis de ambas fracciones de elusión de E2C-11 por electroforesis en gel y por espectrometría de masas mostró únicamente la presencia del monómero de E2C-11 (no mostrado), ratificando que el pico adicional se corresponde con una tetramerización de E2C-11. A su vez, la re-inyección en el tamiz molecular de la fracción correspondiente al tetrámero presentó un solo pico correspondiente al peso molecular del dímero (no mostrado). Por lo tanto,

a diferencia de los dominios E2C-16, E2C-18, E2C-31 y E2C-45, la proteína E2C-11 puede tetramerizar en condiciones extremas de concentración, en equilibrio con el estado dimérico (**Figura 5A**). Esta es la primer evidencia de tetramerización del dominio E2C, aunque justamente para E2C-11 se ha reportado recientemente en estudios de microscopía electrónica, la propensión del dominio de formar multímeros estando unida a sitios de ADN específico (Sim *et al.* 2008), y además la facilidad de la proteína ‘entera’ E2 de HPV11, de hexamerizar vía el extremo N-terminal al formar el complejo con los sitios de ADN BS1, BS2 y BS3 (Sim *et al.* 2008). La oligomerización de E2 vía el dominio N-terminal resulta clave a la hora de formar bucles en el genoma viral y permitir regulaciones génicas de remota distancia (Hernandez-Ramon *et al.* 2008), de facilitar la replicación viral (Sim *et al.* 2008) o de favorecer la segregación del genoma de HPV con el genoma del huésped (McBride 2008).

Más allá de la función poco explorada de la oligomerización de E2 inducida por la unión en simultáneo a distintos sitios específicos de ADN, es evidente la importancia del evento como herramienta adicional de regulación del ciclo infeccioso del virus (Hernandez-Ramon *et al.* 2008; McBride 2008; Sim *et al.* 2008). En nuestro estudio biofísico podemos asignar el origen entálpico observado en el reconocimiento cooperativo de E2C-11 (**Figura V.6**), a esta facilidad para la tetramerización del dominio (**Figura 5A**), en línea con evidencias físicas reportadas (Sim *et al.* 2008). Además, los cambios en la conformación del DBS inducidos por el dominio E2C-11 al DBS fueron más lentos que los cambios observados para los otros dos dominios (**Figura V.9**), una diferencia más que podría estar también relacionada con las diferencias termodinámicas observadas. El comportamiento para los dominios E2C-16 y 18 es opuesto, su energía favorable para la unión al DBS es la entropía (**Figura V.6**), los cambios conformacionales ocurren en forma concertada con la unión al DBS y no hemos observado ni se han reportado en la literatura eventos de tetramerización.

Finalmente, estudios adicionales a los presentados en la sección resultados de la presente tesis muestran diferencias contundentes en la desnaturalización térmica de los dominios E2C-11, 16 y 18. Los dominios E2C-11 y 18 presentan una desnaturalización reversible (**Figura 5B** y **5C**, respectivamente), resultando menos estable el dominio E2C-11 que el dominio E2C-18, en línea con lo observado en el Capítulo IV. Este comportamiento térmico contrasta rotundamente con lo observado para E2C-16, que agrega en forma desordenada como estudiamos en el primer Capítulo de la tesis (**Figura I.1**). En tanto el dominio 31 y el 45 presentaron desnaturalizaciones

térmicas irreversibles (no mostrado). Estas diferencias pueden también estar vinculadas con las diferencias observadas en el reconocimiento cooperativo, pues probablemente involucran regiones proteicas expuestas al solvente no conservadas entre los dominios. Por otro lado inducen a pensar que el mecanismo de agregación en E2C-11 y E2C-18 es diferente y experimentos futuros de polimerización entre varios dominios homólogos E2C pueden ser claves para definir con exactitud la secuencia amiloidogénica del dímero.

La activación de la replicación del ADN está asociada con el programa de transcripción. En el caso de virus de ADN doble cadena, la replicación del genoma reprime la expresión de proteínas tempranas. Esto incluye aquellas participando directamente en la replicación, como el caso de E2 en HPV. Consecuentemente, se desencadena la expresión de proteínas tardías, incluyendo los elementos de la cápside, la coraza que luego transportará el ADN amplificado a nuevos huéspedes. Nuestra caracterización estructural y termodinámica de un evento de unión clave para el acoplamiento entre la replicación y la transcripción en HPV puede ayudar a comprender este proceso en virus de ADN doble cadena y en sitios adyacentes en general.

Epílogo.

De acuerdo al diccionario de Oxford, el adjetivo *proteico* significa: “que tiende o puede cambiar con frecuencia o fácilmente”; “capaz de hacer muchas cosas diferentes”. La etimología de *proteico* es Proteo, anciano Dios del mar en la mitología griega. El profético Proteo se caracteriza por modificar su apariencia con la intención de esconderse y evitar predecir el futuro a aquellos que se lo requieran. Los sinónimos de *proteico* se apoyan en esta particularidad del Dios del mar: adaptable, cambiante, mudable, mutable, voluble, veleidoso, versátil, variable, vacilante, inconstante, inconsistente, inestable, fluctuante, flexible, líquido, lábil y polifacético. Así como también se establecen los antónimos: inmutable e invariable. La ilustración de la estructura tridimensional de una proteína, se aproxima más su antónimo que a su significado. El pintor surrealista René Magritte ilustra este concepto en su obra maestra La traición de las imágenes. El cuadro al óleo consiste en la imagen de una pipa y la inscripción: “*Ceci n’est pas une pipe*”. Burllescamente el autor expresa que la representación de una pipa no es una pipa. Del mismo modo, la representación tridimensional de una proteína le escapa a su definición. Rindiendo un humilde homenaje a René Magritte: “*Esto no es una proteína*”.

En el presente trabajo de tesis se estudiaron algunas aristas de la multifuncional proteína

E2 de papilomavirus humano. Su inestabilidad conformacional desemboca en polifacéticos agregados solubles de tipo amiloideo. Su flexible interacción con ADN, su adaptable estructura dimérica, lábil a mínimas vacilaciones en su entorno, y su versátil comportamiento sinérgico convergen en una fina regulación del ciclo infeccioso viral coordinada con la maquinaria del huésped.



Tesista

Lic. Mariano Dellarole



Director de Tesis

Dr. Gonzalo de Prat Gay

Quinta Sección

Referencias

*In science as in other fields of endeavor, one finds saints and charlatans,
warriors and monks, geniuses and cranks,
tyrants and slaves, benefactors and misers,
but there is one quality that the best of them have in common,
one that they share with great writers,
musicians, and artists:
creativity.*

Max F. Perutz, Is Science Necessary?

- Abbate, E. A., C. Voitenleitner and M. R. Botchan (2006). "Structure of the papillomavirus DNA-tethering complex E2:Brd4 and a peptide that ablates HPV chromosomal association." Mol Cell **24**(6): 877-89.
- Alexander, K. A. and W. C. Phelps (1996). "A fluorescence anisotropy study of DNA binding by HPV-11 E2C protein: a hierarchy of E2-binding sites." Biochemistry **35**(30): 9864-72.
- Alexandrov, B. S., V. Gelev, A. R. Bishop, A. Usheva and K. O. Rasmussen (2010). "DNA Breathing Dynamics in the Presence of a Terahertz Field." Phys Lett A **374**(10): 1214.
- Alexandrov, B. S., V. Gelev, S. W. Yoo, L. B. Alexandrov, Y. Fukuyo, A. R. Bishop, K. O. Rasmussen and A. Usheva (2010). "DNA dynamics play a role as a basal transcription factor in the positioning and regulation of gene transcription initiation." Nucleic Acids Res **38**(6): 1790-5.
- Alonso, L. G., C. Smal, M. M. Garcia-Alai, L. Chemes, M. Salame and G. d. Prat Gay (2006). "Chaperone holdase activity of human papillomavirus E7 oncoprotein." Biochemistry **45**(3): 657-67.
- Arber, W. (1965). "Host-controlled modification of bacteriophage." Annu Rev Microbiol **19**: 365-78.
- Ashkenazy, H., E. Erez, E. Martz, T. Pupko and N. Ben-Tal (2010). "ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids." Nucleic Acids Res **38** Suppl: W529-33.
- Ausbel, F., R. Brent, R. E. Kingston, D. D. Moore, J. G. Siedman, J. G. Smith and K. Struhl (1997). Short protocols in molecular biology. New York.
- Baase, W. A. and W. C. Johnson, Jr. (1979). "Circular dichroism and DNA secondary structure." Nucleic Acids Res **6**(2): 797-814.
- Badal, S., V. Badal, I. E. Calleja-Macias, M. Kalantari, L. S. Chuang, B. F. Li and H. U. Bernard (2004). "The human papillomavirus-18 genome is efficiently targeted by cellular DNA methylation." Virology **324**(2): 483-92.
- Badal, V., L. S. Chuang, E. H. Tan, S. Badal, L. L. Villa, C. M. Wheeler, B. F. Li and H. U. Bernard (2003). "CpG methylation of human papillomavirus type 16 DNA in cervical cancer cell lines and in clinical specimens: genomic hypomethylation correlates with carcinogenic progression." J Virol **77**(11): 6227-34.
- Bains, G. and E. Freire (1991). "Calorimetric determination of cooperative interactions in high affinity binding processes." Anal Biochem **192**(1): 203-6.
- Baker, B. M. and K. P. Murphy (1996). "Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry." Biophys J **71**(4): 2049-55.

- Ballin, J. D., I. A. Shkel and M. T. Record, Jr. (2004). "Interactions of the KWK6 cationic peptide with short nucleic acid oligomers: demonstration of large Coulombic end effects on binding at 0.1-0.2 M salt." Nucleic Acids Res **32**(11): 3271-81.
- Barbic, A., D. P. Zimmer and D. M. Crothers (2003). "Structural origins of adenine-tract bending." Proc Natl Acad Sci U S A **100**(5): 2369-73.
- Bedrosian, C. L. and D. Bastia (1990). "The DNA-binding domain of HPV-16 E2 protein interaction with the viral enhancer: protein-induced DNA bending and role of the nonconserved core sequence in binding site affinity." Virology **174**(2): 557-75.
- Beechem, J. M. (1992). "Global analysis of biochemical and biophysical data." Methods Enzymol **210**: 37-54.
- Berg, O. G. and P. H. von Hippel (1987). "Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters." J Mol Biol **193**(4): 723-50.
- Berg, O. G. and P. H. von Hippel (1988). "Selection of DNA binding sites by regulatory proteins." Trends Biochem Sci **13**(6): 207-11.
- Berg, O. G. and P. H. von Hippel (1988). "Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites." J Mol Biol **200**(4): 709-23.
- Berg, P., D. Baltimore, S. Brenner, R. O. Roblin and M. F. Singer (1975). "Summary statement of the Asilomar conference on recombinant DNA molecules." Proc Natl Acad Sci U S A **72**(6): 1981-4.
- Berg, P. and J. E. Mertz (2010). "Personal reflections on the origins and emergence of recombinant DNA technology." Genetics **184**(1): 9-17.
- Berg, P. and M. F. Singer (1995). "The recombinant DNA controversy: twenty years later." Proc Natl Acad Sci U S A **92**(20): 9011-3.
- Bernard, H. U. (2006). Phylogeny and taxonomy of papillomaviruses. Papillomavirus Research: From Natural History to Vaccines and Beyond. M. Saveria Campo. Wymondham, Caisreir Academic Press. **Chapter 2**: 11-18.
- Bernard, H. U., R. D. Burk, Z. Chen, K. van Doorslaer, H. Z. Hausen and E. M. de Villiers (2010). "Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments." Virology.
- Beynon, R. (1996). "Buffer Calculator." Java based Molecular Biologist's Workbench, <http://www.bioinformatics.org/JaMBW>.
- Bhattacharjee, B. and S. Sengupta (2006). "CpG methylation of HPV 16 LCR at E2 binding site proximal to P97 is associated with cervical cancer in presence of intact E2." Virology **354**(2): 280-5.

- Biancalana, M. and S. Koide (2010). "Molecular mechanism of Thioflavin-T binding to amyloid fibrils." Biochim Biophys Acta **1804**(7): 1405-1412.
- Biancalana, M., K. Makabe, A. Koide and S. Koide (2009). "Molecular mechanism of thioflavin-T binding to the surface of beta-rich peptide self-assemblies." J Mol Biol **385**(4): 1052-63.
- Bioedit "URL: <http://www.mbio.ncsu.edu/BioEdit/>."
- Blakaj, D. M., C. Kattamuri, S. Khrapunov, R. S. Hegde and M. Brenowitz (2006). "Indirect readout of DNA sequence by papillomavirus E2 proteins depends upon net cation uptake." J Mol Biol **358**(1): 224-40.
- Bochkarev, A., J. A. Barwell, R. A. Pfuetzner, E. Bochkareva, L. Frappier and A. M. Edwards (1996). "Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein, EBNA1, bound to DNA." Cell **84**(5): 791-800.
- Bochkarev, A., J. A. Barwell, R. A. Pfuetzner, W. Furey, Jr., A. M. Edwards and L. Frappier (1995). "Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein EBNA 1." Cell **83**(1): 39-46.
- Bolognesi, B., J. R. Kumita, T. P. Barros, E. K. Esbjorner, L. M. Luheshi, D. C. Crowther, M. R. Wilson, C. M. Dobson, G. Favrin and J. J. Yerbury (2010). "ANS binding reveals common features of cytotoxic amyloid species." ACS Chem Biol **5**(8): 735-40.
- Bosch, F. X., S. Sanjosé, X. Castellsagué, V. Moreno and N. Muñoz (2006). Epimediology of Human Papillomavirus Infections and Associations with Cervical Cancer: New Opportunities for Prevention. Papillomavirus Research: From Natural History to Vaccines and Beyond. M. Saveria Campo. Wymondham, Caisreir Academic Press. **Chapter 3**: 19-40.
- Bose, K., N. C. Yoder, K. Kumar and J. D. Baleja (2007). "The role of conserved histidines in the structure and stability of human papillomavirus type 16 E2 DNA-binding domain." Biochemistry **46**(5): 1402-11.
- Brandts, J. F., H. R. Halvorson and M. Brennan (1975). "Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues." Biochemistry **14**: 4953-4963.
- Bravo, I. G. and A. Alonso (2007). "Phylogeny and evolution of papillomaviruses based on the E1 and E2 proteins." Virus Genes **34**(3): 249-62.
- Brohawn, S. G., J. R. Partridge, J. R. Whittle and T. U. Schwartz (2009). "The nuclear pore complex has entered the atomic age." Structure **17**(9): 1156-68.
- Brown, A. (2009). "Analysis of cooperativity by isothermal titration calorimetry." Int J Mol Sci **10**(8): 3457-77.

- Brown, C., K. Campos-Leon, M. Strickland, C. Williams, V. Fairweather, R. L. Brady, M. P. Crump and K. Gaston (2010). "Protein flexibility directs DNA recognition by the papillomavirus E2 proteins." Nucleic Acids Res.
- Brown, J. H. (2006). "Breaking symmetry in protein dimers: designs and functions." Protein Sci **15**(1): 1-13.
- Büchen-Osmond, C. (2003). "The Universal Virus Database ICTVdB." Computing in Science & Engineering **5** (3): 16-25.
- Buck, M. (1998). "Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins." Q Rev Biophys **31**(3): 297-355.
- Bussiere, D. E., X. Kong, D. A. Egan, K. Walter, T. F. Holzman, F. Lindh, T. Robins and V. L. Giranda (1998). "Structure of the E2 DNA-binding domain from human papillomavirus serotype 31 at 2.4 Å." Acta Crystallogr D Biol Crystallogr **54**(Pt 6 Pt 2): 1367-76.
- Byun, K. S. and D. L. Beveridge (2004). "Molecular dynamics simulations of papilloma virus E2 DNA sequences: dynamical models for oligonucleotide structures in solution." Biopolymers **73**(3): 369-79.
- Chan, S. Y., L. Ho, C. K. Ong, V. Chow, B. Drescher, M. Durst, J. ter Meulen, L. Villa, J. Luande, H. N. Mgaya and et al. (1992). "Molecular variants of human papillomavirus type 16 from four continents suggest ancient pandemic spread of the virus and its coevolution with humankind." J Virol **66**(4): 2057-66.
- Chen, X. S., R. L. Garcea, I. Goldberg, G. Casini and S. C. Harrison (2000). "Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16." Mol Cell **5**(3): 557-67.
- Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins and J. D. Thompson (2003). "Multiple sequence alignment with the Clustal series of programs." Nucleic Acids Res **31**(13): 3497-500.
- Chiang, C. M., G. Dong, T. R. Broker and L. T. Chow (1992). "Control of human papillomavirus type 11 origin of replication by the E2 family of transcription regulatory proteins." J Virol **66**(9): 5224-31.
- Chiang, C. M., M. Ustav, A. Stenlund, T. F. Ho, T. R. Broker and L. T. Chow (1992). "Viral E1 and E2 proteins support replication of homologous and heterologous papillomaviral origins." Proc Natl Acad Sci U S A **89**(13): 5799-803.
- Chiti, F., P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi and C. M. Dobson (1999). "Designing conditions for in vitro formation of amyloid protofilaments and fibrils." Proc Natl Acad Sci U S A **96**(7): 3590-4.

- Cicero, D. O., A. D. Nadra, T. Eliseo, M. Dellarole, M. Paci and G. d. Prat Gay (2006). "Structural and thermodynamic basis for the enhanced transcriptional control by the human papillomavirus strain-16 E2 protein." Biochemistry **45**(21): 6551-60.
- Clamp, M., J. Cuff, S. M. Searle and G. J. Barton (2004). "The Jalview Java alignment editor." Bioinformatics **20**(3): 426-7.
- Cohen, J. (2005). "Public health. High hopes and dilemmas for a cervical cancer vaccine." Science **308**(5722): 618-21.
- Colon, W. (1999). "Analysis of protein structure by solution optical spectroscopy." Methods Enzymol **309**: 605-32.
- Cox, M., K. Ryder, S. Silver and P. Tegtmeyer (1988). "The role of operator position in SV40 T-antigen-mediated repression." Virology **167**(1): 293-5.
- Crick, F. (1970). "Central dogma of molecular biology." Nature **227**(5258): 561-3.
- Crick, F. (1970). "Molecular biology in the year 2000." Nature **228**(5272): 613-5.
- Crick, F. H. (1958). "On protein synthesis." Symp Soc Exp Biol **12**: 138-63.
- Crooks, G. E., G. Hon, J. M. Chandonia and S. E. Brenner (2004). "WebLogo: a sequence logo generator." Genome Res **14**(6): 1188-90.
- Crothers, D. M. (1998). "DNA curvature and deformation in protein-DNA complexes: a step in the right direction." Proc Natl Acad Sci U S A **95**(26): 15163-5.
- Crothers, D. M., T. E. Haran and J. G. Nadeau (1990). "Intrinsically bent DNA." J Biol Chem **265**(13): 7093-6.
- D.R., H. P. M. a. L. (2001). Papillomaviruses and their Replication. Fields Virology. K. D. M. a. H. P.M. Philadelphia, Lippincott Williams and Wilkins. **1**: 2213-2219.
- Dantur, K., L. Alonso, E. Castano, L. Morelli, J. M. Centeno-Crowley, S. Vighi and G. de Prat-Gay (2009). "Cytosolic accumulation of HPV16 E7 oligomers supports different transformation routes for the prototypic viral oncoprotein: the amyloid-cancer connection." Int J Cancer **125**(8): 1902-11.
- de Villiers, E. M., C. Fauquet, T. R. Broker, H. U. Bernard and H. zur Hausen (2004). "Classification of papillomaviruses." Virology **324**(1): 17-27.
- de Villiers, E. M. and K. Gunst (2009). "Characterization of seven novel human papillomavirus types isolated from cutaneous tissue, but also present in mucosal lesions." J Gen Virol **90**(Pt 8): 1999-2004.
- Dell, G., K. W. Wilkinson, R. Tranter, J. Parish, R. Leo Brady and K. Gaston (2003). "Comparison of the structure and DNA-binding properties of the E2 proteins from an oncogenic and a non-oncogenic human papillomavirus." J Mol Biol **334**(5): 979-91.

- Dellarole, M. (2005). Diseño y caracterización de variantes monoméricas del dominio dimérico de unión a ADN del factor de transcripción E2 del Papilomavirus Humano., Universidad de Buenos Aires.
- Dellarole, M., I. E. Sanchez and G. de Prat Gay (2010). "Thermodynamics of Cooperative DNA Recognition at a Replication Origin and Transcription Regulatory Site." Biochemistry **49**(48): 10277-10286.
- Dellarole, M., I. E. Sanchez, E. Freire and G. d. Prat Gay (2007). "Increased Stability and DNA Site Discrimination of "Single Chain" Variants of the Dimeric beta-Barrel DNA Binding Domain of the Human Papillomavirus E2 Transcriptional Regulator." Biochemistry **46**(43): 12441-50.
- Demeret, C., C. Desaintes, M. Yaniv and F. Thierry (1997). "Different mechanisms contribute to the E2-mediated transcriptional repression of human papillomavirus type 18 viral oncogenes." Journal of Virology **71**(12): 9343-9.
- Demeret, C., M. Le Moal, M. Yaniv and F. Thierry (1995). "Control of HPV 18 DNA replication by cellular and viral transcription factors." Nucleic Acids Res **23**(23): 4777-84.
- Di Pietro, S. M., J. M. Centeno, M. L. Cerutti, M. F. Lodeiro, D. U. Ferreira, L. G. Alonso, F. P. Schwarz, F. A. Goldbaum and G. d. Prat Gay (2003). "Specific antibody-DNA interaction: a novel strategy for tight DNA recognition." Biochemistry **42**(20): 6218-27.
- Djuranovic, D., C. Oguey and B. Hartmann (2004). "The role of DNA structure and dynamics in the recognition of bovine papillomavirus E2 protein target sequences." J Mol Biol **339**(4): 785-96.
- Dobson, C. M. and M. Karplus (1999). "The fundamentals of protein folding: bringing together theory and experiment." Curr Opin Struct Biol **9**(1): 92-101.
- Dong, G., T. R. Broker and L. T. Chow (1994). "Human papillomavirus type 11 E2 proteins repress the homologous E6 promoter by interfering with the binding of host transcription factors to adjacent elements." J Virol **68**(2): 1115-27.
- Doorbar, J. (2006). "Molecular biology of human papillomavirus infection and cervical cancer." Clin Sci (Lond) **110**(5): 525-41.
- Douglas, S. E. (1995). "DNA Strider. An inexpensive sequence analysis package for the Macintosh." Mol Biotechnol **3**(1): 37-45.
- Dragan, A. I., R. Carrillo, T. I. Gerasimova and P. L. Privalov (2008). "Assembling the human IFN-beta enhanceosome in solution." J Mol Biol **384**(2): 335-48.
- Edelhoch, H. (1967). "Spectroscopic determination of tryptophan and tyrosine in proteins." Biochemistry **6**: 1948-1954.
- Eftink, M. R., A. C. Anusiem and R. L. Biltonen (1983). "Enthalpy-entropy compensation and heat capacity changes for protein-ligand interactions: general thermodynamic models and data for the binding of nucleotides to ribonuclease A." Biochemistry **22**(16): 3884-96.

- Eliseo, T., I. E. Sanchez, A. D. Nadra, M. Dellarole, M. Paci, G. de Prat Gay and D. O. Cicero (2009). "Indirect DNA readout on the protein side: coupling between histidine protonation, global structural cooperativity, dynamics, and DNA binding of the human papillomavirus type 16 E2C domain." J Mol Biol **388**(2): 327-44.
- Ellis, K. J. and J. F. Morrison (1982). "Buffers of constant ionic strength for studying pH-dependent processes." Methods Enzymol **87**: 405-26.
- Ensser, A. and H. Pfister (1990). "Epidermodysplasia verruciformis associated human papillomaviruses present a subgenus-specific organization of the regulatory genome region." Nucleic Acids Res **18**(13): 3919-22.
- Falconi, M., F. Oteri, T. Eliseo, D. O. Cicero and A. Desideri (2008). "MD simulations of papillomavirus DNA-E2 protein complexes hints at a protein structural code for DNA deformation." Biophys J **95**(3): 1108-17.
- Falconi, M., A. Santolamazza, T. Eliseo, G. d. Prat Gay, D. O. Cicero and A. Desideri (2007). "Molecular dynamics of the DNA-binding domain of the papillomavirus E2 transcriptional regulator uncover differential properties for DNA target accommodation." Febs J **274**(9): 2385-95.
- Fandrich, M. (2007). "Absolute correlation between lag time and growth rate in the spontaneous formation of several amyloid-like aggregates and fibrils." J Mol Biol **365**(5): 1266-70.
- Feeney, K. M. and J. L. Parish (2009). "Targeting mitotic chromosomes: a conserved mechanism to ensure viral genome persistence." Proc Biol Sci **276**(1662): 1535-44.
- Ferreiro, D. U. (2003). Mecanismo de unión a ADN del dominio C-terminal de la proteína E2 del Papilomavirus Humano, principal reglador de la expresión génica viral., Universidad de Buenos Aires.
- Ferreiro, D. U., M. Dellarole, A. D. Nadra and G. d. Prat Gay (2005). "Free energy contributions to direct readout of a DNA sequence." J Biol Chem **280**(37): 32480-4.
- Ferreiro, D. U., L. M. Lima, A. D. Nadra, L. G. Alonso, F. A. Goldbaum and G. d. Prat Gay (2000). "Distinctive cognate sequence discrimination, bound DNA conformation, and binding modes in the E2 C-terminal domains from prototype human and bovine papillomaviruses." Biochemistry **39**(47): 14692-701.
- Ferreiro, D. U. and G. d. Prat Gay (2003). "A protein-DNA binding mechanism proceeds through multi-state or two-state parallel pathways." J Mol Biol **331**(1): 89-99.
- Ferreiro, D. U., I. E. Sanchez and G. de Prat Gay (2008). "Transition state for protein-DNA recognition." Proc Natl Acad Sci U S A **105**(31): 10797-802.
- Ferrone, F. (1999). "Analysis of protein aggregation kinetics." Methods Enzymol **309**: 256-74.

- Flint, S. J., L. W. Enquist, V. R. Racaniello and A. M. Skalka (2008). Principles of Virology. Principles of Virology, 3rd Edition BUNDLE, ASM Press. **1**: 288-333.
- Flint, S. J., L. W. Enquist, V. R. Racaniello and A. M. Skalka (2008). Transcription strategies: DNA templates Principles of Virology, 3rd Edition BUNDLE, ASM Press. **1**: 240-287.
- Foguel, D., J. L. Silva and G. d. Prat Gay (1998). "Characterization of a partially folded monomer of the DNA-binding domain of human papillomavirus E2 protein obtained at high pressure." J Biol Chem **273**(15): 9050-7.
- Fontaine, V., E. van der Meijden, J. de Graaf, J. ter Schegget and L. Struyk (2000). "A functional NF-kappaB binding site in the human papillomavirus type 16 long control region." Virology **272**(1): 40-9.
- Forslund, O., T. Iftner, K. Andersson, B. Lindelof, E. Hradil, P. Nordin, B. Stenquist, R. Kirnbauer, J. Dillner and E. M. de Villiers (2007). "Cutaneous human papillomaviruses found in sun-exposed skin: Beta-papillomavirus species 2 predominates in squamous cell carcinoma." J Infect Dis **196**(6): 876-83.
- Fowler, D. M., A. V. Koulov, W. E. Balch and J. W. Kelly (2007). "Functional amyloid--from bacteria to humans." Trends Biochem Sci **32**(5): 217-24.
- Freire, E., C. Oddo, L. Frappier and G. de Prat-Gay (2008). "Kinetically driven refolding of the hyperstable EBNA1 origin DNA-binding dimeric beta-barrel domain into amyloid-like spherical oligomers." Proteins **70**(2): 450-61.
- Freire, E., A. Schon and A. Velazquez-Campoy (2009). "Isothermal titration calorimetry: general formalism using binding polynomials." Methods Enzymol **455**: 127-55.
- Garber, A. C., J. Hu and R. Renne (2002). "Latency-associated nuclear antigen (LANA) cooperatively binds to two sites within the terminal repeat, and both sites contribute to the ability of LANA to suppress transcription and to facilitate DNA replication." J Biol Chem **277**(30): 27401-11.
- Garcia-Alai, M. M., L. G. Alonso and G. d. Prat Gay (2007). "The N-terminal module of HPV16 E7 is an intrinsically disordered domain that confers conformational and recognition plasticity to the oncoprotein." Biochemistry **46**(37): 10405-12.
- Garcia-Vallve, S., A. Alonso and I. G. Bravo (2005). "Papillomaviruses: different genes have different histories." Trends Microbiol **13**(11): 514-21.
- Garcia-Vallve, S., J. R. Iglesias-Rozas, A. Alonso and I. G. Bravo (2006). "Different papillomaviruses have different repertoires of transcription factor binding sites: convergence and divergence in the upstream regulatory region." BMC Evol Biol **6**: 20.

- Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch (2003). "ExPASy: The proteomics server for in-depth protein knowledge and analysis." Nucleic Acids Res **31**(13): 3784-8.
- Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel and A. Bairoch (2005). Protein identification and analysis tools on the ExPASy Server. The Proteomics Protocols Handbook. J. M. Walker, Humana Press: 571-607.
- Gloss, B., H. U. Bernard, K. Seedorf and G. Klock (1987). "The upstream regulatory region of the human papilloma virus-16 contains an E2 protein-independent enhancer which is specific for cervical carcinoma cells and regulated by glucocorticoid hormones." Embo J **6**(12): 3735-43.
- Goldstein, R. F. and L. Stryer (1986). "Cooperative polymerization reactions. Analytical approximations, numerical examples, and experimental strategy." Biophys J **50**(4): 583-99.
- Gray, D. (1996). Circular Dichroism and the Conformational Analysis of Biomolecules. New York, Plenum Press.
- Gray, D. (1996). Circular Dichroism of Protein nucleic Acid interactions. Circular Dichroism and the Conformational Analysis of Biomolecules. G. Fasman. New York, Plenum Press.
- Gray, D. M., S. H. Hung and K. H. Johnson (1995). "Absorption and circular dichroism spectroscopy of nucleic acid duplexes and triplexes." Methods Enzymol **246**: 19-34.
- Greene, R. F., Jr. and C. N. Pace (1974). "Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin." J Biol Chem **249**(17): 5388-93.
- Guido, M. C., R. Zamorano, E. Garrido-Guerrero, P. Gariglio and A. Garcia-Carranca (1992). "Early promoters of genital and cutaneous human papillomaviruses are differentially regulated by the bovine papillomavirus type 1 E2 gene product." J Gen Virol **73 (Pt 6)**: 1395-400.
- Gunasekaran, K., S. J. Eyles, A. T. Hagler and L. M. Gierasch (2001). "Keeping it in the family: folding studies of related proteins." Curr. Opin. Struct. Biol. **11**(1): 83-93.
- Ha, J. H., M. W. Capp, M. D. Hohenwalter, M. Baskerville and M. T. Record, Jr. (1992). "Thermodynamic stoichiometries of participation of water, cations and anions in specific and non-specific binding of lac repressor to DNA. Possible thermodynamic origins of the "glutamate effect" on protein-DNA interactions." J Mol Biol **228**(1): 252-64.
- Hall, T. (2005). BioEdit. I. Caredata.com. Carlsbad, CA, Ibis Therapeutics: Biological sequence alignment editor written for Windows 95/98/NT/2000/XP.
- Heer, A., L. G. Alonso and G. de Prat-Gay (2011). "E6*, the 50 Amino Acid Product of the Most Abundant Spliced Transcript of the E6 Oncoprotein in High-Risk Human Papillomavirus, Is a Promiscuous Folder and Binder." Biochemistry.

- Hegde, R. S. (2002). "The papillomavirus E2 proteins: structure, function, and biology." Annu Rev Biophys Biomol Struct **31**: 343-60.
- Hegde, R. S. and E. J. Androphy (1998). "Crystal structure of the E2 DNA-binding domain from human papillomavirus type 16: implications for its DNA binding-site selection mechanism." J Mol Biol **284**(5): 1479-89.
- Hegde, R. S., S. R. Grossman, L. A. Laimins and P. B. Sigler (1992). "Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target." Nature **359**(6395): 505-12.
- Hegde, R. S., A. F. Wang, S. S. Kim and M. Schapira (1998). "Subunit rearrangement accompanies sequence-specific DNA binding by the bovine papillomavirus-1 E2 protein." J Mol Biol **276**(4): 797-808.
- Herbst, L. H., J. Lenz, K. Van Doorslaer, Z. Chen, B. A. Stacy, J. F. Wellehan, Jr., C. A. Manire and R. D. Burk (2009). "Genomic characterization of two novel reptilian papillomaviruses, *Chelonia mydas* papillomavirus 1 and *Caretta caretta* papillomavirus 1." Virology **383**(1): 131-5.
- Hernandez-Ramon, E. E., J. E. Burns, W. Zhang, H. F. Walker, S. Allen, A. A. Antson and N. J. Maitland (2008). "Dimerization of the human papillomavirus type 16 E2 N terminus results in DNA looping within the upstream regulatory region." J Virol **82**(10): 4853-61.
- Hines, C. S., C. Meghoo, S. Shetty, M. Biburger, M. Brenowitz and R. S. Hegde (1998). "DNA structure and flexibility in the sequence-specific binding of papillomavirus E2 proteins." J Mol Biol **276**(4): 809-18.
- Hizver, J., H. Rozenberg, F. Frolow, D. Rabinovich and Z. Shakked (2001). "DNA bending by an adenine--thymine tract and its role in gene regulation." Proc Natl Acad Sci U S A **98**(15): 8490-5.
- Hooley, E., V. Fairweather, A. R. Clarke, K. Gaston and R. L. Brady (2006). "The recognition of local DNA conformation by the human papillomavirus type 6 E2 protein." Nucleic Acids Res **34**(14): 3897-908.
- Horcas, I., R. Fernandez, J. M. Gomez-Rodriguez, J. Colchero, J. Gomez-Herrero and A. M. Baro (2007). "WSXM: a software for scanning probe microscopy and a tool for nanotechnology." Rev Sci Instrum **78**(1): 013705.
- Hou, S. Y., S. Y. Wu and C. M. Chiang (2002). "Transcriptional activity among high and low risk human papillomavirus E2 proteins correlates with E2 DNA binding." J Biol Chem **277**(47): 45619-29.
- Houtman, J. C., P. H. Brown, B. Bowden, H. Yamaguchi, E. Appella, L. E. Samelson and P. Schuck (2007). "Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling." Protein Sci **16**(1): 30-42.

- Howley, P. (2001). Papillomaviruses and their replication. Fields Virology. D. Knipe and P. Howley, ed. Philadelphia, Lippincot Williams and Wilkins.
- Hoyer, W., D. Cherny, V. Subramaniam and T. M. Jovin (2004). "Rapid self-assembly of alpha-synuclein observed by in situ atomic force microscopy." J Mol Biol **340**(1): 127-39.
- IARC, W. G. (2007). Monograph on Human Papillomaviruses, World Health Organization, INTERNATIONAL AGENCY FOR RESEARCH ON CANCER. **90-6**.
- Ikai, A. (1971). The kinetics of unfolding and refolding of proteins. Department of Biochemistry, Duke University.
- Ishimaru, D., L. R. Andrade, L. S. Teixeira, P. A. Quesado, L. M. Maiolino, P. M. Lopez, Y. Cordeiro, L. T. Costa, W. M. Heckl, G. Weissmuller, D. Foguel and J. L. Silva (2003). "Fibrillar aggregates of the tumor suppressor p53 core domain." Biochemistry **42**(30): 9022-7.
- Ishimaru, D., A. P. Ano Bom, L. M. Lima, P. A. Quesado, M. F. Oyama, C. V. de Moura Gallo, Y. Cordeiro and J. L. Silva (2009). "Cognate DNA stabilizes the tumor suppressor p53 and prevents misfolding and aggregation." Biochemistry **48**(26): 6126-35.
- Jackson, D. A., R. H. Symons and P. Berg (1972). "Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli." Proc Natl Acad Sci U S A **69**(10): 2904-9.
- Janin, J. (1999). "Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition." Structure Fold Des **7**(12): R277-9.
- Jen-Jacobson, L., L. E. Engler and L. A. Jacobson (2000). "Structural and thermodynamic strategies for site-specific DNA binding proteins." Structure Fold Des **8**(10): 1015-23.
- Jha, S. K. and J. B. Udgaonkar (2009). "Direct evidence for a dry molten globule intermediate during the unfolding of a small protein." Proc Natl Acad Sci U S A **106**(30): 12289-94.
- Johnson, B. B., K. S. Dahl, I. Tinoco, Jr., V. I. Ivanov and V. B. Zhurkin (1981). "Correlations between deoxyribonucleic acid structural parameters and calculated circular dichroism spectra." Biochemistry **20**(1): 73-8.
- Kalantari, M. and H.-U. Bernard (2006). Gene Expression of Papillomaviruses. Papillomavirus Research: From Natural History to Vaccines and Beyond. M. Saveria Campo. Wymondham, Caisreir Academic Press. **Chapter 4**: 41-52.
- Kalantari, M., I. E. Calleja-Macias, D. Tewari, B. Hagmar, K. Lie, H. A. Barrera-Saldana, D. J. Wiley and H. U. Bernard (2004). "Conserved methylation patterns of human papillomavirus type 16 DNA in asymptomatic infection and cervical neoplasia." J Virol **78**(23): 12762-72.

- Kalodimos, C. G., A. M. Bonvin, R. K. Salinas, R. Wechselberger, R. Boelens and R. Kaptein (2002). "Plasticity in protein-DNA recognition: lac repressor interacts with its natural operator O1 through alternative conformations of its DNA-binding domain." Embo J **21**(12): 2866-76.
- Kelly, J. W. (1998). "The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways." Curr Opin Struct Biol **8**(1): 101-6.
- Kelly, T. J. (1988). "SV40 DNA replication." J Biol Chem **263**(34): 17889-92.
- Kiefhaber, T., I. E. Sanchez and A. Bachmann (2008). Characterization of Protein Folding Barriers with Rate Equilibrium Free Energy Relationships. Protein Folding Handbook. J. B. a. T. Kiefhaber, Wiley-VCH. **1**: 411-44.
- Kim, K., P. A. Garner-Hamrick, C. Fisher, D. Lee and P. F. Lambert (2003). "Methylation patterns of papillomavirus DNA, its influence on E2 function, and implications in viral infection." J Virol **77**(23): 12450-9.
- Kim, S. S., J. K. Tam, A. F. Wang and R. S. Hegde (2000). "The structural basis of DNA target discrimination by papillomavirus E2 proteins." J Biol Chem **275**(40): 31245-54.
- King, D. A., L. Zhang, L. Guarente and R. Marmorstein (1999). "Structure of a HAP1-DNA complex reveals dramatically asymmetric DNA binding by a homodimeric protein." Nat Struct Biol **6**(1): 64-71.
- Kirillov, A., B. Kistler, R. Mostoslavsky, H. Cedar, T. Wirth and Y. Bergman (1996). "A role for nuclear NF-kappaB in B-cell-specific demethylation of the Igkappa locus." Nat Genet **13**(4): 435-41.
- Klucevsek, K., M. Wertz, J. Lucchi, A. Leszczynski and J. Moroianu (2007). "Characterization of the nuclear localization signal of high risk HPV16 E2 protein." Virology **360**(1): 191-8.
- Klunk, W. E., R. F. Jacob and R. P. Mason (1999). "Quantifying amyloid by congo red spectral shift assay." Methods Enzymol **309**: 285-305.
- Kodali, R. and R. Wetzel (2007). "Polymorphism in the intermediates and products of amyloid assembly." Curr Opin Struct Biol **17**(1): 48-57.
- Komatsu, T., M. E. Ballestas, A. J. Barbera, B. Kelley-Clarke and K. M. Kaye (2004). "KSHV LANA1 binds DNA as an oligomer and residues N-terminal to the oligomerization domain are essential for DNA binding, replication, and episome persistence." Virology **319**(2): 225-36.
- Kumar, S. and J. B. Udgaonkar (2009). "Conformational conversion may precede or follow aggregate elongation on alternative pathways of amyloid protofibril formation." J Mol Biol **385**(4): 1266-76.
- Kumar, S. and J. B. Udgaonkar (2009). "Structurally distinct amyloid protofibrils form on separate pathways of aggregation of a small protein." Biochemistry **48**(27): 6441-9.

- Lacal, J., M. E. Guazzaroni, P. Gutierrez-del-Arroyo, A. Busch, M. Velez, T. Krell and J. L. Ramos (2008). "Two levels of cooperativeness in the binding of TodT to the tod operon promoter." J Mol Biol **384**(5): 1037-47.
- Lakowicz, J. R. (1983). Principles of Fluorescence Spectroscopy. New York, Plenum Press.
- Lakowicz, J. R. (1999). Principles of fluorescence spectroscopy. New York.
- Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-8.
- Laskowski, R. A. (2009). "PDBsum new things." Nucleic Acids Res **37**(Database issue): D355-9.
- Leffler, J. E. and E. Grunwald (1963). Rates and equilibria of organic reactions. New York, Dover.
- Lendel, C., B. Bolognesi, A. Wahlstrom, C. M. Dobson and A. Graslund (2010). "Detergent-like interaction of Congo red with the amyloid beta peptide." Biochemistry **49**(7): 1358-60.
- LeVine, H., 3rd (1999). "Quantification of beta-sheet amyloid fibril structures with thioflavin T." Methods Enzymol **309**: 274-84.
- Liang, H., A. M. Petros, R. P. Meadows, H. S. Yoon, D. A. Egan, K. Walter, T. F. Holzman, T. Robins and S. W. Fesik (1996). "Solution structure of the DNA-binding domain of a human papillomavirus E2 protein: evidence for flexible DNA-binding regions." Biochemistry **35**(7): 2095-103.
- Lim, C., H. Sohn, Y. Gwack and J. Choe (2000). "Latency-associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus (human herpesvirus-8) binds ATF4/CREB2 and inhibits its transcriptional activation activity." J Gen Virol **81**(Pt 11): 2645-52.
- Lima, L. M., D. Foguel and J. L. Silva (2000). "DNA tightens the dimeric DNA-binding domain of human papillomavirus E2 protein without changes in volume." Proc Natl Acad Sci U S A **97**(26): 14289-94.
- Lima, L. M. and G. d. Prat Gay (1997). "Conformational changes and stabilization induced by ligand binding in the DNA-binding domain of the E2 protein from human papillomavirus." J Biol Chem **272**(31): 19295-303.
- Lima, L. M. and J. L. Silva (2004). "Positive contribution of hydration on DNA binding by E2c protein from papillomavirus." J Biol Chem **279**(46): 47968-74.
- Liu, B. and B. M. Alberts (1995). "Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex." Science **267**(5201): 1131-7.
- Lorenzo, A. and B. A. Yankner (1994). "Beta-amyloid neurotoxicity requires fibril formation and is inhibited by congo red." Proc Natl Acad Sci U S A **91**(25): 12243-7.

- Lu, J. Z., Y. N. Sun, R. C. Rose, W. Bonneze and D. J. McCance (1993). "Two E2 binding sites (E2BS) alone or one E2BS plus an A/T-rich region are minimal requirements for the replication of the human papillomavirus type 11 origin." J Virol **67**(12): 7131-9.
- Luheshi, L. M., G. G. Tartaglia, A. C. Brorsson, A. P. Pawar, I. E. Watson, F. Chiti, M. Vendruscolo, D. A. Lomas, C. M. Dobson and D. C. Crowther (2007). "Systematic in vivo analysis of the intrinsic determinants of amyloid Beta pathogenicity." PLoS Biol **5**(11): e290.
- MacDonald, D., K. Herbert, X. Zhang, T. Pologruto, P. Lu and T. Polgruto (2001). "Solution structure of an A-tract DNA bend." J Mol Biol **306**(5): 1081-98.
- MacPhee, C. E. and C. M. Dobson (2000). "Chemical dissection and reassembly of amyloid fibrils formed by a peptide fragment of transthyretin." J Mol Biol **297**(5): 1203-15.
- MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, University of California Press. **1**: 281-297.
- Maji, S. K., M. H. Perrin, M. R. Sawaya, S. Jessberger, K. Vadodaria, R. A. Rissman, P. S. Singru, K. P. Nilsson, R. Simon, D. Schubert, D. Eisenberg, J. Rivier, P. Sawchenko, W. Vale and R. Riek (2009). "Functional amyloids as natural storage of peptide hormones in pituitary secretory granules." Science **325**(5938): 328-32.
- Marcon, G., G. Plakoutsi and F. Chiti (2006). "Protein aggregation starting from the native globular state." Methods Enzymol **413**: 75-91.
- Matsuo, K., J. Silke, O. Georgiev, P. Marti, N. Giovannini and D. Rungger (1998). "An embryonic demethylation mechanism involving binding of transcription factors to replicating DNA." Embo J **17**(5): 1446-53.
- Matthews, C. R. (1987). "Effect of point mutations on the folding of globular proteins." Meth. Enzymol. **154**: 498-511.
- McBride, A. A. (2008). "Replication and partitioning of papillomavirus genomes." Adv Virus Res **72**: 155-205.
- McBride, A. A., H. Romanczuk and P. M. Howley (1991). "The papillomavirus E2 regulatory proteins." J Biol Chem **266**(28): 18411-4.
- McNeil, C. (2006). "Who invented the VLP cervical cancer vaccines?" J Natl Cancer Inst **98**(7): 433.
- Modis, Y., B. L. Trus and S. C. Harrison (2002). "Atomic model of the papillomavirus capsid." Embo J **21**(18): 4754-62.
- Mohr, I. J., R. Clark, S. Sun, E. J. Androphy, P. MacPherson and M. R. Botchan (1990). "Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator." Science **250**(4988): 1694-9.

- Mok, Y. K. (1996). Studies on the DNA binding domain from human papillomavirus strain 16 E2 protein. Cambridge, University of Cambridge.
- Mok, Y. K., L. G. Alonso, L. M. Lima, M. Bycroft and G. d. Prat Gay (2000). "Folding of a dimeric beta-barrel: residual structure in the urea denatured state of the human papillomavirus E2 DNA binding domain." Protein Sci **9**(4): 799-811.
- Mok, Y. K., M. Bycroft and G. d. Prat Gay (1996). "The dimeric DNA binding domain of the human papillomavirus E2 protein folds through a monomeric intermediate which cannot be native-like." Nat Struct Biol **3**(8): 711-7.
- Mok, Y. K., G. d. Prat Gay, P. J. Butler and M. Bycroft (1996). "Equilibrium dissociation and unfolding of the dimeric human papillomavirus strain-16 E2 DNA-binding domain." Protein Sci **5**(2): 310-9.
- Morrow, J. F., S. N. Cohen, A. C. Chang, H. W. Boyer, H. M. Goodman and R. B. Helling (1974). "Replication and transcription of eukaryotic DNA in Escherichia coli." Proc Natl Acad Sci U S A **71**(5): 1743-7.
- Muller, F., T. Giroglou and M. Sapp (1997). "Characterization of the DNA-binding activity of the E1 and E2 proteins and the E1/E2 complex of human papillomavirus type 33." J Gen Virol **78** (Pt 4): 911-5.
- Munoz, N. (2000). "Human papillomavirus and cancer: the epidemiological evidence." J Clin Virol **19**(1-2): 1-5.
- Munoz, N., F. X. Bosch, S. de Sanjose, R. Herrero, X. Castellsague, K. V. Shah, P. J. Snijders and C. J. Meijer (2003). "Epidemiologic classification of human papillomavirus types associated with cervical cancer." N Engl J Med **348**(6): 518-27.
- Musheev, M. U., Y. Filiptsev, V. Okhonin and S. N. Krylov (2010). "Electric Field Destabilizes Noncovalent Protein-DNA Complexes." J Am Chem Soc.
- Myers, J. K., C. N. Pace and J. M. Scholtz (1995). "Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding." Protein Sci **4**(10): 2138-48.
- Nadra, A. D. (2005). Estudios estructurales en solución del factor de Transcripción de E2C de HPV-16 y su interacción con ADN (PhD thesis). Instituto Leloir. Buenos Aires, Universidad de Buenos Aires.
- Nadra, A. D., T. Eliseo, Y. K. Mok, C. L. Almeida, M. Bycroft, M. Paci, G. d. Prat Gay and D. O. Cicero (2004). "Solution structure of the HPV-16 E2 DNA binding domain, a transcriptional regulator with a dimeric beta-barrel fold." J Biomol NMR **30**(2): 211-4.

- Nakamura, T., K. Makabe, K. Tomoyori, K. Maki, A. Mukaiyama and K. Kuwajima (2010). "Different folding pathways taken by highly homologous proteins, goat alpha-lactalbumin and canine milk lysozyme." J Mol Biol **396**(5): 1361-78.
- Newhouse, C. D. and S. J. Silverstein (2001). "Orientation of a novel DNA binding site affects human papillomavirus-mediated transcription and replication." J Virol **75**(4): 1722-35.
- Nickson, A. A. and J. Clarke (2010). "What lessons can be learned from studying the folding of homologous proteins?" Methods **52**(1): 38-50.
- O'Brien, R., B. DeDecker, K. G. Fleming, P. B. Sigler and J. E. Ladbury (1998). "The effects of salt on the TATA binding protein-DNA interaction from a hyperthermophilic archaeon." J Mol Biol **279**(1): 117-25.
- O'Nuallain, B., A. K. Thakur, A. D. Williams, A. M. Bhattacharyya, S. Chen, G. Thiagarajan and R. Wetzel (2006). "Kinetics and thermodynamics of amyloid assembly using a high-performance liquid chromatography-based sedimentation assay." Methods Enzymol **413**: 34-74.
- Orth, G. (1999). "Human papillomaviruses: general features." Clin Dermatol **15**(2): 181-98.
- Pace, C. N. (1986). "Determination and analysis of urea and guanidine hydrochloride denaturation curves." Methods Enzymol **131**: 266-80.
- Pace, C. N., F. Vajdos, F. Lanette, G. Grimsley and T. Gray (1995). "How to measure and predict the molar absorption coefficient of a protein." Protein Science **4**: 2411-23.
- Pallares, I., J. Vendrell, F. X. Aviles and S. Ventura (2004). "Amyloid fibril formation by a partially structured intermediate state of alpha-chymotrypsin." J Mol Biol **342**(1): 321-31.
- Papanicolaou, G. N. (1942). "A New Procedure for Staining Vaginal Smears." Science **95**(2469): 438-9.
- Parish, J. L., A. Kowalczyk, H. T. Chen, G. E. Roeder, R. Sessions, M. Buckle and K. Gaston (2006). "E2 proteins from high- and low-risk human papillomavirus types differ in their ability to bind p53 and induce apoptotic cell death." J Virol **80**(9): 4580-90.
- Peng, J. and J. Xu (2010). "Low-homology protein threading." Bioinformatics **26**(12): i294-300.
- Perez-Bercoff, A., J. Koch and T. R. Burglin (2006). "LogoBar: bar graph visualization of protein logos with gaps." Bioinformatics **22**(1): 112-4.
- Polvino-Bodnar, M. and P. A. Schaffer (1992). "DNA binding activity is required for EBNA 1-dependent transcriptional activation and DNA replication." Virology **187**(2): 591-603.
- Powers, E. T. and D. L. Powers (2006). "The kinetics of nucleated polymerizations at high concentrations: amyloid fibril formation near and above the "supercritical concentration"." Biophys J **91**(1): 122-32.
- Prado, F. and A. Aguilera (2005). "Impairment of replication fork progression mediates RNA polIII transcription-associated recombination." Embo J **24**(6): 1267-76.

- Prat Gay, G. d., K. Gaston and D. O. Cicero (2008). "The papillomavirus E2 DNA binding domain." Front Biosci **13**: 6006-21.
- Prat Gay, G. d., A. D. Nadra, F. J. Corrales-Izquierdo, L. G. Alonso, D. U. Ferreira and Y. K. Mok (2005). "The folding mechanism of a dimeric beta-barrel domain." J Mol Biol **351**(3): 672-82.
- Ptashne, M. (2004). A Genetic Switch - Phage Lambda Revisited. Cold Spring Harbor, Cold Spring Harbor Laboratory Press.
- Pyles, E. A. and J. C. Lee (1998). "Escherichia coli cAMP receptor protein-DNA complexes. 2. Structural asymmetry of DNA bending." Biochemistry **37**(15): 5201-10.
- Ramirez-Alvarado, M., J. S. Merkel and L. Regan (2000). "A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro." Proc Natl Acad Sci U S A **97**(16): 8979-84.
- Ramirez-Carrozzi, V. and T. Kerppola (2003). "Asymmetric recognition of nonconsensus AP-1 sites by Fos-Jun and Jun-Jun influences transcriptional cooperativity with NFAT1." Mol Cell Biol **23**(5): 1737-49.
- Rapp, B., A. Pawellek, F. Kraetzer, M. Schaefer, C. May, K. Purdie, K. Grassmann and T. Iftner (1997). "Cell-type-specific separate regulation of the E6 and E7 promoters of human papillomavirus type 6a by the viral transcription factor E2." Journal of Virology **71**(9): 6956-66.
- Raumann, B. E., M. A. Rould, C. O. Pabo and R. T. Sauer (1994). "DNA recognition by beta-sheets in the Arc repressor-operator crystal structure." Nature **367**(6465): 754-7.
- Rawlins, D. R., G. Milman, S. D. Hayward and G. S. Hayward (1985). "Sequence-specific DNA binding of the Epstein-Barr virus nuclear antigen (EBNA-1) to clustered sites in the plasmid maintenance region." Cell **42**(3): 859-68.
- Record, M. T., Jr., C. F. Anderson and T. M. Lohman (1978). "Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity." Q Rev Biophys **11**(2): 103-78.
- Record, M. T., Jr., J. H. Ha and M. A. Fisher (1991). "Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA." Methods Enzymol **208**: 291-343.
- Reeck, G. R., C. de Haen, D. C. Teller, R. F. Doolittle, W. M. Fitch, R. E. Dickerson, P. Chambon, A. D. McLachlan, E. Margoliash, T. H. Jukes and et al. (1987). "'Homology" in proteins and nucleic acids: a terminology muddle and a way out of it." Cell **50**(5): 667.
- Remm, M., R. Brain and J. R. Jenkins (1992). "The E2 binding sites determine the efficiency of replication for the origin of human papillomavirus type 18." Nucleic Acids Res **20**(22): 6015-21.

- Rivals, I., L. Personnaz, L. Taing and M. C. Potier (2007). "Enrichment or depletion of a GO category within a class of genes: which test?" Bioinformatics **23**(4): 401-7.
- Robbins, J. B., M. C. Murphy, B. A. White, R. I. Mackie, T. Ha and I. K. Cann (2004). "Functional analysis of multiple single-stranded DNA-binding proteins from *Methanosarcina acetivorans* and their effects on DNA synthesis by DNA polymerase BI." J Biol Chem **279**(8): 6315-26.
- Rohs, R., H. Sklenar and Z. Shakked (2005). "Structural and energetic origins of sequence-specific DNA bending: Monte Carlo simulations of papillomavirus E2-DNA binding sites." Structure **13**(10): 1499-509.
- Rosl, F., A. Arab, B. Klevenz and H. zur Hausen (1993). "The effect of DNA methylation on gene regulation of human papillomaviruses." J Gen Virol **74** (Pt 5): 791-801.
- Rozenberg, H., D. Rabinovich, F. Frolow, R. S. Hegde and Z. Shakked (1998). "Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets." Proc Natl Acad Sci U S A **95**(26): 15194-9.
- Sanchez, I. E., M. Dellarole, K. Gaston and G. de Prat Gay (2008). "Comprehensive comparison of the interaction of the E2 master regulator with its cognate target DNA sites in 73 human papillomavirus types by sequence statistics." Nucleic Acids Res **36**(3): 756-69.
- Sanchez, I. E., D. U. Ferreira, M. Dellarole and G. d. P. Gay (2010). "Experimental snapshots of a protein-DNA binding landscape." Proceedings of the National Academy of Sciences of the USA.
- Sanders, C. M. and N. J. Maitland (1994). "Kinetic and equilibrium binding studies of the human papillomavirus type-16 transcription regulatory protein E2 interacting with core enhancer elements." Nucleic Acids Res **22**(23): 4890-7.
- Santoro, M. M. and D. W. Bolen (1988). "Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants." Biochemistry **27**(21): 8063-8.
- Sarai, A. and Y. Takeda (1989). "Lambda repressor recognizes the approximately 2-fold symmetric half-operator sequences asymmetrically." Proc Natl Acad Sci U S A **86**(17): 6513-7.
- Schiffman, M. and P. E. Castle (2003). "Human papillomavirus: epidemiology and public health." Arch Pathol Lab Med **127**(8): 930-4.
- Schneider, T. D. and R. M. Stephens (1990). "Sequence logos: a new way to display consensus sequences." Nucleic Acids Res **18**(20): 6097-100.
- Schumacher, M. A., M. C. Miller, S. Grkovic, M. H. Brown, R. A. Skurray and R. G. Brennan (2002). "Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR." Embo J **21**(5): 1210-8.

- Sekhar, A. and J. B. Udgaonkar (2011). "Fluoroalcohol-Induced Modulation of the Pathway of Amyloid Protofibril Formation by Barstar." Biochemistry.
- Semisotnov, G. V., N. A. Rodionova, O. I. Razgulyaev, V. N. Uversky, A. F. Gripas and R. I. Gilmanshin (1991). "Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe." Biopolymers **31**(1): 119-28.
- Shatzky-Schwartz, M., N. D. Arbuckle, M. Eisenstein, D. Rabinovich, A. Bareket-Samish, T. E. Haran, B. F. Luisi and Z. Shakked (1997). "X-ray and solution studies of DNA oligomers and implications for the structural basis of A-tract-dependent curvature." J Mol Biol **267**(3): 595-623.
- Sichero, L. and L. L. Villa (2006). "Epidemiological and functional implications of molecular variants of human papillomavirus." Braz J Med Biol Res **39**(6): 707-17.
- Silva, J. L., T. C. Vieira, M. P. Gomes, A. P. Bom, L. M. Lima, M. S. Freitas, D. Ishimaru, Y. Cordeiro and D. Foguel (2010). "Ligand binding and hydration in protein misfolding: insights from studies of prion and p53 tumor suppressor proteins." Acc Chem Res **43**(2): 271-9.
- Sim, J., S. Ozgur, B. Y. Lin, J. H. Yu, T. R. Broker, L. T. Chow and J. Griffith (2008). "Remodeling of the human papillomavirus type 11 replication origin into discrete nucleoprotein particles and looped structures by the E2 protein." J Mol Biol **375**(4): 1165-77.
- Smal, C., D. E. Wetzler, K. I. Dantur, L. B. Chemes, M. M. Garcia-Alai, M. Dellarole, L. G. Alonso, K. Gaston and G. de Prat-Gay (2009). "The human papillomavirus E7-E2 interaction mechanism in vitro reveals a finely tuned system for modulating available E7 and E2 proteins." Biochemistry **48**(50): 11939-49.
- Soeda, E., M. C. Ferran, C. C. Baker and A. A. McBride (2006). "Repression of HPV16 early region transcription by the E2 protein." Virology.
- Spalholz, B. A., J. C. Byrne and P. M. Howley (1988). "Evidence for cooperativity between E2 binding sites in E2 trans-regulation of bovine papillomavirus type 1." J Virol **62**(9): 3143-50.
- Srisailam, S., T. K. Kumar, D. Rajalingam, K. M. Kathir, H. S. Sheu, F. J. Jan, P. C. Chao and C. Yu (2003). "Amyloid-like fibril formation in an all beta-barrel protein. Partially structured intermediate state(s) is a precursor for fibril formation." J Biol Chem **278**(20): 17701-9.
- Steger, G. and S. Corbach (1997). "Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein." Journal of Virology **71**(1): 50-8.
- Steger, G., M. Jarzabek-Chorzelska, S. Jablonska and H. Pfister (1988). "Human papillomavirus 8-L1 immune serum: a new diagnostic possibility for Epidermodysplasia verruciformis-specific HPVs." J Invest Dermatol **91**(1): 76-81.

- Stubenrauch, F., I. M. Leigh and H. Pfister (1996). "E2 represses the late gene promoter of human papillomavirus type 8 at high concentrations by interfering with cellular factors." Journal of Virology **70**(1): 119-26.
- Stubenrauch, F., H. B. Lim and L. A. Laimins (1998). "Differential requirements for conserved E2 binding sites in the life cycle of oncogenic human papillomavirus type 31." Journal of Virology **72**(2): 1071-7.
- Summers, H., J. A. Barwell, R. A. Pfuetzner, A. M. Edwards and L. Frappier (1996). "Cooperative assembly of EBNA1 on the Epstein-Barr virus latent origin of replication." J Virol **70**(2): 1228-31.
- Sverdrup, F. and S. A. Khan (1995). "Two E2 binding sites alone are sufficient to function as the minimal origin of replication of human papillomavirus type 18 DNA." J Virol **69**(2): 1319-23.
- Tachezy, R., A. Rector, M. Havelkova, E. Wollants, P. Fiten, G. Opdenakker, B. Jenson, J. Sundberg and M. Van Ranst (2002). "Avian papillomaviruses: the parrot *Psittacus erithacus* papillomavirus (PePV) genome has a unique organization of the early protein region and is phylogenetically related to the chaffinch papillomavirus." BMC Microbiol **2**: 19.
- Tan, S., L. E. Leong, P. A. Walker and H. Bernard (1994). "The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID." Journal of Virology **68**(10): 6411-20.
- Tan, S. H., B. Gloss and H. U. Bernard (1992). "During negative regulation of the human papillomavirus-16 E6 promoter, the viral E2 protein can displace Sp1 from a proximal promoter element." Nucleic Acids Research **20**(2): 251-6.
- Tanford, C. (1968). "Protein denaturation. Part B. The transition from native to denatured state." Adv. Prot. Chem **23**: 218-282.
- Tanford, C. (1970). "Protein Denaturation. Part C. Theoretical models for the mechanism of denaturation." Adv. Prot. Chem. **24**: 1-95.
- Tatusova, T. A. and T. L. Madden (1999). "BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences." FEMS Microbiol Lett **174**(2): 247-50.
- Thain, A., O. Jenkins, A. R. Clarke and K. Gaston (1996). "CpG methylation directly inhibits binding of the human papillomavirus type 16 E2 protein to specific DNA sequences." Journal of Virology **70**(10): 7233-5.
- Thain, A., K. Webster, D. Emery, A. R. Clarke and K. Gaston (1997). "DNA binding and bending by the human papillomavirus type 16 E2 protein. Recognition of an extended binding site." J Biol Chem **272**(13): 8236-42.

- Thierry, F. (2009). "Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma." Virology **384**(2): 375-9.
- Thierry, F. and P. M. Howley (1991). "Functional analysis of E2-mediated repression of the HPV18 P105 promoter." New Biol **3**(1): 90-100.
- Thomassin, H., M. Flavin, M. L. Espinas and T. Grange (2001). "Glucocorticoid-induced DNA demethylation and gene memory during development." Embo J **20**(8): 1974-83.
- Tippmann, H. F. (2004). "Analysis for free: comparing programs for sequence analysis." Brief Bioinform **5**(1): 82-7.
- Titolo, S., K. Brault, J. Majewski, P. W. White and J. Archambault (2003). "Characterization of the minimal DNA binding domain of the human papillomavirus e1 helicase: fluorescence anisotropy studies and characterization of a dimerization-defective mutant protein." J Virol **77**(9): 5178-91.
- Titolo, S., E. Welchner, P. W. White and J. Archambault (2003). "Characterization of the DNA-binding properties of the origin-binding domain of simian virus 40 large T antigen by fluorescence anisotropy." J Virol **77**(9): 5512-8.
- Townson, S. A., J. C. Samuelson, Y. Bao, S. Y. Xu and A. K. Aggarwal (2007). "BstYI bound to noncognate DNA reveals a "hemispecific" complex: implications for DNA scanning." Structure **15**(4): 449-59.
- Ullrich, A., J. Shine, J. Chirgwin, R. Pictet, E. Tischer, W. J. Rutter and H. M. Goodman (1977). "Rat insulin genes: construction of plasmids containing the coding sequences." Science **196**(4296): 1313-9.
- Vacic, V., L. M. Iakoucheva and P. Radivojac (2006). "Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments." Bioinformatics **22**(12): 1536-7.
- van Heel, M. (1984). "Multivariate statistical classification of noisy images (randomly oriented biological macromolecules)." Ultramicroscopy **13**(1-2): 165-83.
- van Heel, M., G. Harauz, E. V. Orlova, R. Schmidt and M. Schatz (1996). "A new generation of the IMAGIC image processing system." J Struct Biol **116**(1): 17-24.
- Vassar, P. S. and C. F. Culling (1959). "Fluorescent stains, with special reference to amyloid and connective tissues." Arch Pathol **68**: 487-98.
- Veeraraghavan, S., C. C. Mello, E. J. Androphy and J. D. Baleja (1999). "Structural correlates for enhanced stability in the E2 DNA-binding domain from bovine papillomavirus." Biochemistry **38**(49): 16115-24.
- Weinberg, R. L., D. B. Veprintsev and A. R. Fersht (2004). "Cooperative binding of tetrameric p53 to DNA." J Mol Biol **341**(5): 1145-59.

- Weisshart, K., P. Taneja, A. Jenne, U. Herbig, D. T. Simmons and E. Fanning (1999). "Two regions of simian virus 40 T antigen determine cooperativity of double-hexamer assembly on the viral origin of DNA replication and promote hexamer interactions during bidirectional origin DNA unwinding." *J Virol* **73**(3): 2201-11.
- Wetzler, D. E., E. M. Castano and G. d. Prat Gay (2007). "A quasi-spontaneous amyloid route in a DNA binding gene regulatory domain: The papillomavirus HPV16 E2 protein." *Protein Sci* **16**(4): 744-54.
- WHO (2007). HPV and Cervical Cancer in the World S. d. S. X. Castellsagué, T. Aguado, K.S. Louie, L. Bruni, J. Muñoz, M. Diaz, K. Irwin, M. Gacic, O. Beauvais, G. Albero, E. Ferrer, S. Byrne, F.X. Bosch World Health Organization: 1:241.
- Wiley, D. J., J. Huh, J. Y. Rao, C. Chang, M. Goetz, M. Poulter, E. Masongsong, C. I. Chang and H. U. Bernard (2005). "Methylation of human papillomavirus genomes in cells of anal epithelia of HIV-infected men." *J Acquir Immune Defic Syndr* **39**(2): 143-51.
- Wilson, V. G., M. West, K. Woytek and D. Rangasamy (2002). "Papillomavirus E1 proteins: form, function, and features." *Virus Genes* **24**(3): 275-90.
- Winston, R. L., J. A. Ehley, E. E. Baird, P. B. Dervan and J. M. Gottesfeld (2000). "Asymmetric DNA binding by a homodimeric bHLH protein." *Biochemistry* **39**(31): 9092-8.
- Wiseman, T., S. Williston, J. F. Brandts and L. N. Lin (1989). "Rapid measurement of binding constants and heats of binding using a new titration calorimeter." *Anal Biochem* **179**(1): 131-7.
- Wong, L. Y. and A. C. Wilson (2005). "Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen induces a strong bend on binding to terminal repeat DNA." *J Virol* **79**(21): 13829-36.
- Wood, C. E., S. H. Tannehill-Gregg, Z. Chen, K. van Doorslaer, D. R. Nelson, J. M. Cline and R. D. Burk (2010). "Novel Betapapillomavirus Associated With Hand and Foot Papillomas in a Cynomolgus Macaque." *Vet Pathol*.
- Workman, C. T., Y. Yin, D. L. Corcoran, T. Ideker, G. D. Stormo and P. V. Benos (2005). "enoLOGOS: a versatile web tool for energy normalized sequence logos." *Nucleic Acids Res* **33**(Web Server issue): W389-92.
- Wu, M. H., J. Y. Chan, P. Y. Liu, S. T. Liu and S. M. Huang (2007). "Human papillomavirus E2 protein associates with nuclear receptors to stimulate nuclear receptor- and E2-dependent transcriptional activations in human cervical carcinoma cells." *Int J Biochem Cell Biol* **39**(2): 413-25.
- Wyman, J. and S. J. Gill (1990). Binding and linkage: functional chemistry of biological macromolecules. Mill Valley, CA, UNSB.

- Xi, Z., Y. Zhang, R. S. Hegde, Z. Shakked and D. M. Crothers (2010). "Anomalous DNA binding by E2 regulatory protein driven by spacer sequence TATA." Nucleic Acids Res **38**(11): 3827-33.
- Zhang, Y., Z. Xi, R. S. Hegde, Z. Shakked and D. M. Crothers (2004). "Predicting indirect readout effects in protein-DNA interactions." Proc Natl Acad Sci U S A **101**(22): 8337-41.
- Zheng, Z. M. and C. C. Baker (2006). "Papillomavirus genome structure, expression, and post-transcriptional regulation." Front Biosci **11**: 2286-302.
- Zimmerman, J. M. and L. J. Maher, 3rd (2003). "Solution measurement of DNA curvature in papillomavirus E2 binding sites." Nucleic Acids Res **31**(17): 5134-9.
- zur Hausen, H. (1996). "Papillomavirus infections--a major cause of human cancers." Biochim Biophys Acta **1288**(2): F55-78.