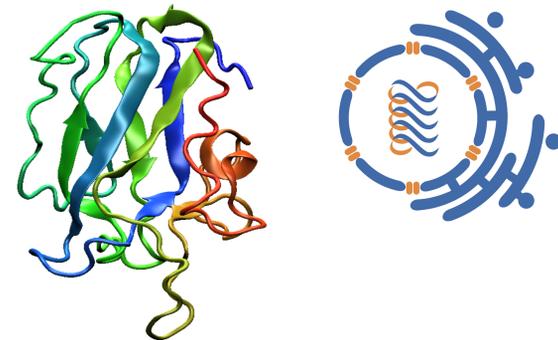


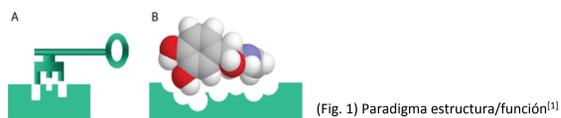
# Comunicación estructural y dinámica de las regiones intrínsecamente desordenadas de la proteína adenoviral E1B-156R.

Marco Antonio Ramírez Martínez  
Asesor: Dra. Carmen Nina Pastor Colón



## ¿Qué son las IDPs?

Las proteínas intrínsecamente desordenadas (IDPs), son proteínas que carecen de una estructura tridimensional definida. Sin embargo éstas con capaces de interactuar con múltiples moléculas y llevar a cabo su función (Fig. 1)<sup>[1]</sup>. Están enriquecidas en aminoácidos que promueven el desorden o rompen con la estructura secundaria de la proteínas, tales como A, R, G, Q, S, P, E, K. Además, han roto el paradigma de estructura/función que postula que se requiere de una estructura definida y fija para tener una función<sup>[1]</sup>.

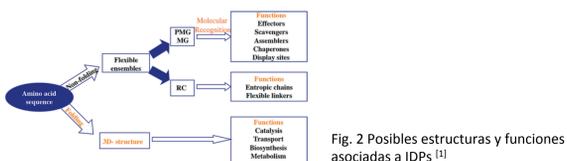


Las IDPs se han estudiado a nivel de estructura primaria mediante:

- Predictores de desorden
- MoRFs (características de reconocimiento molecular, o Molecular Recognition Features)
- Identificadores de motivos de interacción como SIMs (Motivos de interacción con SUMO) y otros.

Todos estos predictores se basan en las propiedades fisicoquímicas de la secuencia o bien, son programas que están entrenados con bases de datos con el fin de mejorar las predicciones<sup>[1]</sup>.

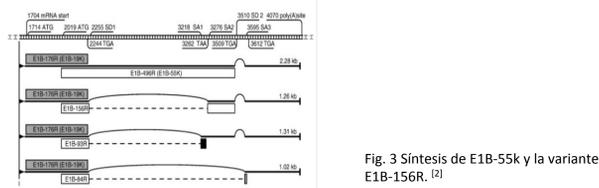
Asociando el grado de desorden que presentan las proteínas se ha podido llegar a la siguiente clasificación de estructuras y funciones (Fig. 2)<sup>[1]</sup>. Las IDPs pueden presentarse como glóbulos fundidos (MG), preglóbulos fundidos (PMG) y estructuras aleatorias (RC).



## ¿Por qué estudiar a E1B-156R?

E1B-156R es una variante de la proteína viral E1B-55K, una proteína multifuncional que inhibe apoptosis, la reparación de fragmentos de ADN, y favorece la traducción de los ARN mensajeros virales. Consta de dos regiones desordenadas, una en el N-terminal y otra en el C-terminal. Su actividad se regula, entre otras, mediante la fosforilación de tres residuos en el C-terminal y esto afecta las modificaciones posttraduccionales en el N-terminal.

E1B-156R es producida por splicing alternativo del transcrito de E1B-55K durante las primeras etapas de infección (Fig. 3). Es de especial interés porque reúne los 79 primeros aminoácidos del N-terminal y los 77 últimos aminoácidos del C-terminal de E1B-55K<sup>[2]</sup>, siendo el modelo mínimo para estudiar las interacciones entre los extremos. E1B-156R participa en el ciclo de infección e incluso puede inhibir apoptosis mediante interacciones con p53 y Daxx<sup>[3]</sup>.



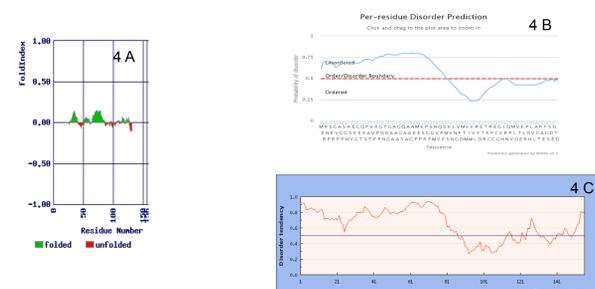
Se han descrito otras variantes de la proteína completa pero no se ha demostrado su papel en infección<sup>[2]</sup>.

Aplicamos herramientas de dinámica molecular para simular a esta proteína y una variante fosforilada, para identificar los cambios provocados por la fosforilación. Elegimos propiedades moleculares que pueden ser medidas experimentalmente, para motivar experimentos futuros.

## Resultados

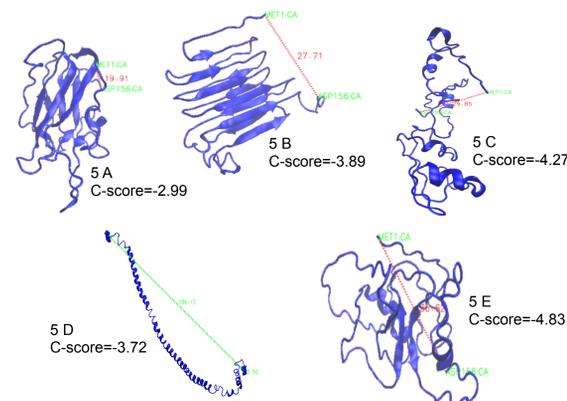
### Predicciones de desorden para E1B-156R

Los resultados arrojados por los predictores (Fig. 4 A FoldIndex<sup>[4]</sup>, B RONNI<sup>[5]</sup>, C DisoPred2<sup>[6]</sup>) nos dieron una idea de las regiones desordenadas presentes en la proteína 156R.



### Modelos tridimensionales para E1B-156R

Obtuvimos modelos tridimensionales utilizando I-Tasser<sup>[7]</sup> (Fig. 5):

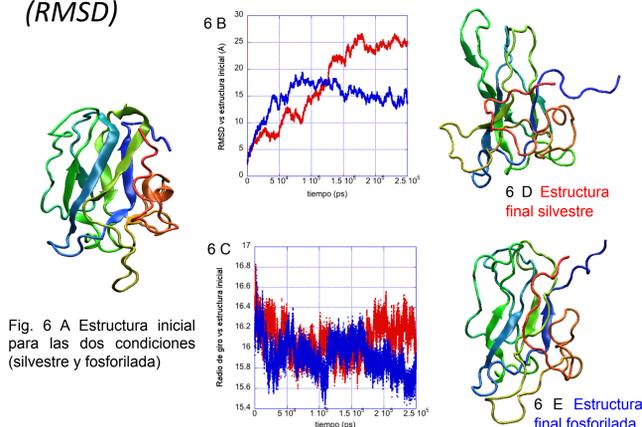


Debido a que nos interesa saber cuales son las interacciones entre los extremos amino y carboxilo terminal, escogimos el modelo con la menor distancia entre los extremos (Fig. 5 A).

Generamos la variante fosforilada en las posiciones 150, 151 y 155 utilizando el programa Charmm-gui<sup>[8,9]</sup>. Con los archivos de coordenadas, potencial y topología generados en este servidor, realizamos 250 ns de simulaciones utilizando NAMD<sup>[10]</sup>. Los resultados se analizaron con Charmm 38<sup>[11]</sup> y Carma<sup>[15]</sup>, y se graficaron con KaleidaGraph<sup>[12]</sup>.

## Resultados de las simulaciones

### Radio de giro y similitud con la estructura inicial (RMSD)



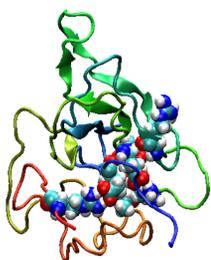
Ambas proteínas partieron de la misma estructura inicial (Fig. 6 A) y se muestran las estructuras finales (Fig. 6 D y 6 E) después de la simulación.

El RMSD (Fig. 6 B) de cada una de las proteínas muestra el cambio de conformación a lo largo de la simulación. Esto nos sugiere que la proteína no ha adoptado todavía una estructura estable, y que además la estructura inicial no es del todo compatible con la secuencia de aminoácidos. El cambio es mayor para la proteína silvestre.

El radio de giro (Fig. 6 B) muestra cómo la proteína fue cambiando de tamaño en el tiempo. Esta propiedad se puede medir en experimentos de dispersión dinámica de luz, filtración en columnas de exclusión molecular y con dispersión de rayos X a bajo ángulo. La proteína fosforilada es un poco más compacta que la silvestre, pero la diferencia es muy pequeña para ser medida.

### Puentes de Hidrógeno

Analizamos las interacciones entre los extremos de las proteínas, y los cambios en poblaciones provocados por la fosforilación.

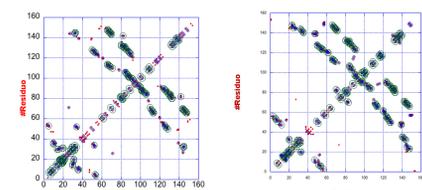


La proteína silvestre (Fig. 7 A) contiene un puente salino que se encarga de unir al aminoácido 3 del amino terminal y al 147 del carboxilo terminal, que de acuerdo con los cálculos, persistió un 76% del tiempo simulado. Además de esos, hay una red de puentes de hidrógeno en una de las caras de la proteína, amarrando la región media de la estructura.



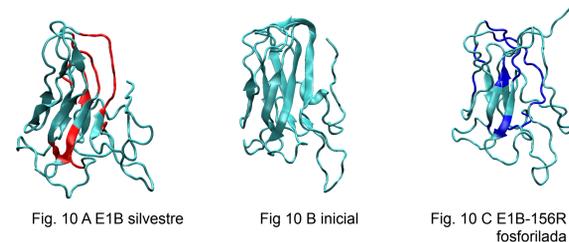
La proteína fosforilada presenta un gran número de puentes de hidrógeno de larga vida (Fig. 7 B). El efecto de la fosforilación está en la Fig. 7 C. El residuo 150 interactúa con cuatro aminoácidos al mismo tiempo, mientras que el 151 se une fuertemente al amino terminal. Estas interacciones no están presentes sin la fosforilación.

### Mapas de Contacto



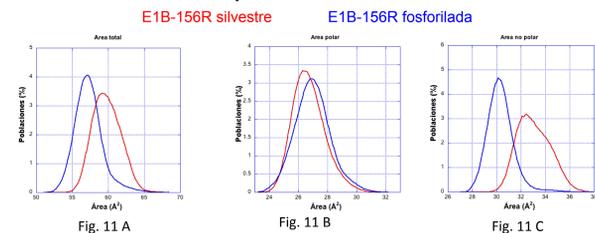
Los mapas de contacto (Fig. 8 A y 8 B) muestran los carbonos que estuvieron a < 6 Å durante el 25% (rojo), 50% (azul) y 75% (verde) del tiempo de simulación. Los mapas de contacto muestran que la arquitectura de ambas proteínas es muy similar. Esto nos dice que el tiempo de simulación es corto aún, y que no han ocurrido todas las consecuencias de haber fosforilado a la proteína.

### Estructura Secundaria



Utilizamos el programa Carma<sup>[15]</sup> para calcular todo tipo de estructura secundaria presente en ambas proteínas (figura 10). Esta propiedad puede compararse con experimentos de dicroísmo circular. Notamos que hay una pérdida de estructura secundaria a lo largo de la simulación. Mostramos la regiones que perdieron estructura (Fig. 10 A y 10 C) con respecto a la estructura inicial (Fig. 10 B).

### Área expuesta al Solvente



Los histogramas muestran que el área total (Fig. 11 A) expuesta es mayor en la versión silvestre. Cuando comparamos las áreas polares (Fig. 11 B) y no polares (Fig. 11 C), podemos observar que la proteína silvestre expone más área no polar. Esto puede ser relacionado con los puentes de hidrógeno en la proteína fosforilada, haciéndola más compacta y por lo tanto con menos superficie no polar expuesta.

## Conclusiones

Nuestra meta fue caracterizar el efecto que tiene la fosforilación a nivel estructural y dinámico en E1B-156R. Los resultados muestran que sí hay diferencias en el comportamiento de las proteínas: la fosforilación promueve compactación mediante la formación de puentes de hidrógeno entre los fosfatos y el N-terminal, y la pérdida de estructura secundaria. Puede ser de interés ver cuál es el efecto que tiene una mutación entre los dos residuos que provocan esta interacción (serina 151 y acilando el N-terminal para eliminar la carga positiva).

Notamos que no hay coincidencias perfectas entre los predictores de desorden, muestra de lo poco que entendemos todavía de las IDPs. Respecto a las dinámicas moleculares, hace falta mucho más tiempo de simulación para comprobar si el core de láminas beta es realmente estable. Otra alternativa es realizar más simulaciones partiendo de las otras estructuras predichas por I-TASSER.

## Agradecimientos

Este proyecto es parte de la Red temática PRODEP "Inmunología molecular de virus". Agradecemos al Centro Nacional de Supercomputo (IPICYT) y al proyecto INFR-2014-02-231504 por tiempo de cómputo. También agradecemos a los alumnos del Laboratorio de Dinámica de Proteínas y Ácidos nucleicos del CIDC de la Universidad Autónoma del Estado de Morelos.

## Referencias

1. Habchi, Tompa. (2014). *Chem. Rev.* 114:6561-6588
2. Sieber et al. (2011). *Virology* 133:133-143
3. Sieber, Dobner. (2007) *J. Virol.* 81:95-105
4. Yang et al (2005). *Bioinformatics.* 21:3369-76.
5. Prilusky et al. (2005). *Bioinformatics* 21:3435-3438.
6. Sickmeier et al. (2007) *Nucleic Acids Res.* 35:86-93
7. Zhang. (2008) *BMC Bioinformatics*.DOI: 10.1186/1471-2105-9-40
8. Jo, et al. (2008) *J. Comput. Chem.* 29:1859-1865
9. Lee, Cheng (2016) *J. Chem. Theory Comput.* 12:405-413
10. Phillips et al (2005). *Journal of Computational Chemistry*, 26:1781-1802.
11. Brooks et. Al (2009). *J. Comput. Chem.* 30:1545-1614
12. KaleidaGraph: Tools for Discovery. 1986-2014 by Synergy Software. Versión 4.5.2
13. Glykos (2006). *Journal of Computational Chemistry*, 27:1765-1768.