

BASE MOLECULAR DEL ACOPLAMIENTO JAZ-MYC, UNA INTERACCIÓN CLAVE EN LA RESPUESTA VEGETAL A ESTRÉS

Miguel Ansel Méndez^{1*}, Sebastian Ayala-Ruano¹, Jonas Goossens², Laurens Pauwels², Alain Goossens², Antonio Leon-Reyes³ y Samara Oña Chuquimarca¹

¹ Grupo de Química Computacional y Teórica, Departamento de Ingeniería Química, Universidad San Francisco de Quito, Quito, Ecuador

² Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

³ Laboratorio de Biotecnología Agrícola y de Alimentos, Ingeniería en Agronomía, Universidad San Francisco de Quito, Quito, Ecuador.

*mmendez@usfq.edu.ec



ABSTRACT

Introducción: La exploración de los mecanismos moleculares de la defensa de plantas nos ha llevado a entender mejor estos sistemas fascinantes, sin embargo la dinámica completa de los procesos de reconocimiento y de acoplamiento proteína-proteína puede profundizarse con el uso de herramientas de modelación y aprendizaje automático. En particular nos interesa entender los mecanismos con que las plantas se adaptan a las amenazas bióticas y del medioambiente. La capacidad vegetal de respuesta ante estrés es de fundamental importancia para la agricultura y la seguridad alimentaria considerando que hay un riesgo real de cambios bruscos en el medio ambiente. El entender cómo las plantas responden al estrés será clave para desarrollos biotecnológicos, y agroquímicos futuros. Los experimentos *in silico* nos permiten a partir de las estructuras moleculares disponibles predecir aquellas que no se han resuelto aun, así como estudiar las interacciones y mecanismos de función con detalle atómico. En particular, reportamos la interacción entre miembros de la familia de proteínas JAZ con un factor de transcripción helice-lazo-helice MYC. Esta interacción es clave en la vía del ácido jasmonico, una molécula pequeña que modula la transcripción de genes relacionados a la respuesta de las plantas a estímulos ambientales [Pauwels and Goossens, 2011].

Resultados y Conclusiones: Se desarrollaron modelos moleculares de todas las proteínas de la familia JAZ (doce proteínas) y de PPD1 y PPD2 también miembros de la familia TIFY. Se desarrolló un flujo de trabajo que hace uso de estas estructuras, y de datos experimentales de mutagénesis puntual que indican la importancia de residuos individuales en la interacción proteína-proteína para mediante la integración de simulaciones de dinámica molecular (50 ns por complejo) y análisis de datos a través de algoritmos de aprendizaje automático mapear todas las proteínas e identificar y cuantificar la importancia de cada residuo (identificación de hot spots) en la interacción proteína-proteína con el factor de transcripción MYC3. Se verificaron las predicciones computacionales mediante la realización de experimentos Y2H (yeast-two-hybrid) donde se observó que la introducción de los aminoácidos clave (para incorporar el motivo SL*FL**R) indujo artificialmente en proteínas PPD la capacidad de interactuar con el factor de transcripción MYC3. Se realizaron mutaciones *in silico* de esos residuos en proteínas PPD1 y PPD2 y la cuantificación teórica de la energía de interacción está en correspondencia con los experimentos de Y2H realizados [Oña, 2020].

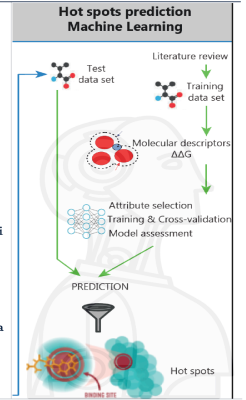
Métodos

La integración de términos energéticos y algoritmos de aprendizaje automático para la predicción de "hot spots" (aminoácidos clave en una interacción proteína - proteína) ha sido explorado por varios autores. Por ejemplo, el cambio de la energía libre de Gibbs luego de una mutación define la clasificación de si un residuo es un hot spot. [Lise et al 2009; Liu et al, 2018]

Pregunta abierta

¿Qué tan efectivo puede ser la implementación de Dinámica molecular (DM) (que generan los términos energéticos) con métodos basados en conocimiento (inteligencia artificial) en obtener predicciones útiles pero rápidas?

Nuestra propuesta: Dinámica molecular full-atom si se puede usar actualmente sin la necesidad de super computadoras pero con HPC si por ejemplo tenemos una familia de proteínas (i.e. TIFY). Con evidencia experimental en uno o pocos de sus miembros servirá para a partir de un número limitado de simulaciones (todos los complejos de interés con los miembros de la familia TIFY con MYC3 (4 simulaciones)). Las configuraciones obtenidas servirán de entrada para CASM y los términos energéticos se calcularán por promedio ponderado del peso de cada configuración.



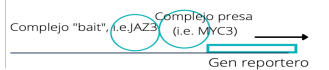
Preparación del set de entrenamiento

- Obtención de descriptores moleculares (valores termodinámicos) de un complejo proteína-proteína:
 - Se analizaron un representante de cada grupo (métodos de agrupación) de las conformaciones en la trayectoria. 2) Se realizó un experimento de Computational alanine scanning mutagenesis (CASM) para cada conformación representativa.
- La "etiqueta" para el entrenamiento fue sacada de la literatura donde una mutación en el residuo X por Y causó una pérdida en la unión entre dos de las proteínas, o no causó una pérdida de la interacción.

Validación in vitro

La secuencia consenso de los resultados para cada par de interacción (hot spots por complejo proteína-proteína(12)), se probó al "retornar" la capacidad de interactuar con MYC3 a dos proteínas que perdieron esta capacidad en algun momento evolutivo: PPD1 y PPD2.

Y2H (yeast two hybrid):

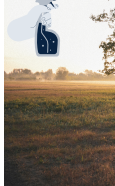


Introducción

Defensa de las plantas

Constitutivo

Estructural
Químico



Mecanismo inducido

Activado solo como respuesta a un ataque o stress
Ejemplo: vía del ácido jasmonico

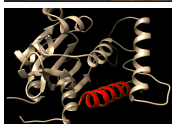
Organismo de estudio

- Arabidopsis thaliana*
- Planta modelo ampliamente usada para estudios en plantas



Objetivo

El principal objetivo es identificar las bases moleculares del acoplamiento JAZn-MYC



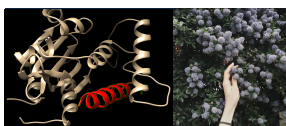
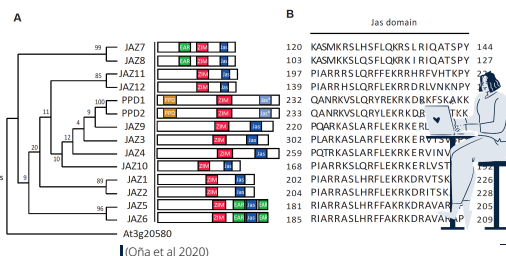
JAZn-MYC

Estructura modelada. En dorado el dominio JID-TAD del factor de transcripción bHLH MYC3. En rojo el dominio JAZ de la proteína represora miembro de la familia clase 2 TIFY.

Proteínas

INTERACCION PROTEINA-PROTEINA

A. Arbol filogenético en *A. Thaliana*. [preparado en MEGA 7]
B. Secuencias del dominio JAS de las proteínas de la familia TIFY clase 2



- ¿Cuáles amino ácidos aportan más a la energía de unión entre las dos proteínas?
- ¿Estos amino ácidos clave (hot spots) son los mismos para toda la familia TIFY?



$$\Delta G_{\text{binding}} = \Delta G_{\text{AB}} - (\Delta G_{\text{A}} + \Delta G_{\text{B}})$$

$$\Delta G_{\text{x}} = a \cdot \Delta G_{\text{vdw}} + b \Delta G_{\text{solvH}} + c \cdot \Delta G_{\text{solvP}} + d \cdot \Delta G_{\text{vd}} + e \cdot \Delta G_{\text{hbOnd}} + f \cdot \Delta G_{\text{el}} + g \cdot \Delta G_{\text{kon}} + h \cdot T \Delta S_{\text{mc}} + k \cdot T \Delta S_{\text{sc}} + l \cdot T \Delta G_{\text{dash}}$$

Resultados y Discusión

Table 3| Summary of the energy profile of JAZ-MYC3, PPD-MYC3 and mPPD-MYC3 complexes.

| Complex | #Residues | Binding free Energy (kcal/mol) | Van der Waals (kcal/mol) | Electrostatics (kcal/mol) | Polar solvation (kcal/mol) | Apolar solvation (kcal/mol) |
|-----------|-----------|--------------------------------|--------------------------|---------------------------|----------------------------|-----------------------------|
| JAZ1-MYC3 | 9 | -1.78 +/- 0.001 | -1.12 +/- 0.011 | -4.22 +/- 0.001 | 15.37 +/- 0.029 | -18.33 +/- 0.014 |
| JAZ2-MYC3 | 11 | -7.09 +/- 0.016 | -13.29 +/- 0.028 | -4.53 +/- 0.007 | 18.04 +/- 0.018 | -17.67 +/- 0.008 |
| JAZ3-MYC3 | 10 | -10.46 +/- 0.011 | -14.09 +/- 0.009 | -3.44 +/- 0.006 | 17.07 +/- 0.013 | -18.61 +/- 0.013 |

Tabla. Parte del set para entrenamiento y validación cruzada usados en la generación de los modelos de aprendizaje automático supervisado (Oña et al, 2020)

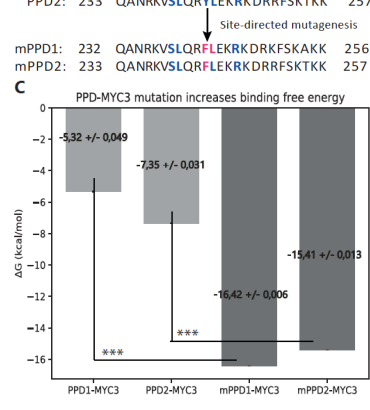
| Category | Residue | Type of Residue | Reference | In-silico site directed mutagenesis results (ΔG, kJ/mol) | | | | | | | |
|----------|---------|-----------------|--|--|-------|------|-------|------|-------|--|--|
| | | | | A | I | K | P | S | V | | |
| ND | M3-92 | TRP | Zhang et al., 2015 | 1.3 | 0.55 | 0.36 | 0.53 | 1.32 | 0.31 | | |
| D | M3-94 | ASP | Smolten et al., 2002; Zhang et al., 2015 | 1.53 | 1.22 | 1.88 | 2.1 | 1.52 | 1.78 | | |
| ND | M3-97 | TYR | Zhang et al., 2017; Zhang et al., 2015 | 0.59 | -0.38 | 0.4 | -0.52 | 0.72 | -0.42 | | |

Ejemplo de modelo de aprendizaje automático entrenado (Random Forest). Donde NDM (mutaciones no-deletéreas) y DM (mutaciones deletéreas)

Table 4| Quantitative assessment of the best ranked machine learning algorithms for JAZ-MYC3 hot spots prediction.

| Algorithm | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|---------|
| Random Forest | 0.818 | 0.125 | 0.900 | 0.818 | 0.857 | 0.685 | 0.945 | 0.959 | NDM |
| | 0.875 | 0.182 | 0.778 | 0.875 | 0.824 | 0.685 | 0.943 | 0.943 | DM |
| | 0.842 | 0.149 | 0.849 | 0.842 | 0.843 | 0.685 | 0.943 | 0.952 | Average |

Tanto las simulaciones como el método in vitro Y2H mostraron que la "restauración" del motivo lineal inducía artificialmente a PPD1 y PPD2 a reconocer a MYC3



Conclusiones

- La predicción de Hot spots reveló un motivo lineal corto que podría definir la especificidad JAZn-MYC3.
- Se logró integrar términos termodinámicos de delta delta G en base a conformaciones calculadas por dinámica molecular como descriptores para modelos de aprendizaje automático supervisado.
- Pequeñas diferencias en la afinidad de unión con el dominio Jas versus el dominio Jas-like en proteínas PPD se deben a la presencia de residuos específicos, "hotspots", que en este caso influyen profundamente en el reconocimiento proteína-proteína.

Agradecimientos

Al Instituto de Simulación Computacional - USFQ por facilitar el uso del HPC-USFQ donde se realizaron todos los cálculos.

Referencias

- Lise et al. BMC Biomf. 2009, 10,365
- Liu et al. Molecules, 2018,23,2535
- Oña, S. et al. Frontiers in Plant Science, 2020, 11, artículo 1139
- Pauwels, L. and Goossens, A., 2011. Plant Cell. 23, 3089-3100