

**UNIVERSIDAD DE CHILE**

**FACULTAD DE CIENCIAS QUÍMICAS Y FARMACÉUTICAS**



***Evolución del mecanismo de intercambio de segmentos en la subfamilia FoxP de factores de transcripción humanos.***

**Tesis presentada a la Universidad de Chile para optar al grado de Doctor en Bioquímica por:**

***PABLO ANDRÉS VILLALOBOS NAVARRO***

**Director de Tesis: Dr. Jorge Babul Cattán  
Co-director de tesis: Dr. César Ramírez-Sarmiento**

**Santiago-CHILE**

**Junio 2021**

# UNIVERSIDAD DE CHILE

## FACULTAD DE CIENCIAS QUÍMICAS Y FARMACÉUTICAS

### INFORME DE APROBACIÓN DE TESIS DE DOCTORADO

Se informa a la Dirección de la Escuela de Graduados de la Facultad de Ciencias Químicas y Farmacéuticas que la Tesis de Doctor presentada por el candidato

**PABLO ANDRÉS VILLALOBOS NAVARRO**

Ha sido aprobada por la Comisión de Evaluadora de Tesis como requisito para optar al grado de Doctor en Bioquímica, en el examen público rendido el día

---

**Director de Tesis:**

**Dr. Jorge Babul Cattan**

---

**Co-director de Tesis:**

**Dr. César Ramírez-Sarmiento**

---

**Comisión Evaluadora de Tesis:**

**Dra. María Cecilia Rojas**

---

**Dr. Christian A.M. Wilson**

---

**Dr. Danilo González-Nilo**

---

**DEDICADO A MIS PADRES Y HERMANOS.**

## AGRADECIMIENTOS

Sin lugar a duda el camino recorrido en una etapa como esta involucra muchas personas. En este espacio me gustaría mencionar y agradecer a aquellas que se me vienen a la mente en este momento y que no puedo dejar de destacar.

Agradezco a mi tutor Dr. Jorge Babul, por estos más de 10 años trabajando en su laboratorio, por su incalculable aporte a mi formación, sin duda será un recuerdo para toda la vida. A la vez, por permitirme tal como Él, entender la ciencia como una labor colaborativa, en donde no se puede avanzar sin pensar en quienes nos rodean, espero que sigamos gritando juntos: ¡Sin ciencia no hay futuro! A mi cotutor, César Ramírez-Sarmiento, por continuamente desafiarme y estimularme a ser mejor y ofrecerme oportunidades para crecer en mi formación, por lo que siempre habrá un gran recuerdo de su aporte en esta etapa.

A mi querida familia, por soportar esta etapa junto a mí, a mis padres por enseñarme con el ejemplo que todo con esfuerzo y perseverancia se puede. Enfrentar este paso de mi formación, especialmente el final en medio de una pandemia ha sido complejo, su apoyo, compañía y comprensión ha sido un soporte incalculable. Son lo mejor y espero que podamos seguir disfrutando de los mejores asados como solo nosotros sabemos.

A los miembros del laboratorio de Bioquímica y Biología Molecular, el que por muchos años ha sido mi segunda casa. Sin dudas, si tuviera que elegir un lugar donde se mezcle la satisfacción personal por el trabajo bien hecho, ideas desafiantes y espacio para crear, elegiría a este laboratorio mil veces. Al Dr. Ricardo Cabrera, por esas conversaciones con un buen café en la cocina, por siempre creer en mí y apoyarme desde un inicio en el lab. Al Dr. Exequiel Medina, por esas conversaciones en la oficina, la amistad y por su tremendo apoyo tanto científico y emocional durante esta tesis, sus consejo y asistencia ha sido un apoyo importante para finalizar esta etapa. A la Dra. Victoria Guixé, por siempre apoyarnos a todas y todos en el lab, por potenciar nuestras ideas, darnos un espacio también para compartir entre profesores y alumnos. Al Dr. Víctor Castro, por también brindarme apoyo siempre que lo requerí, por considerarme en cada viaje en que se pudo ir a difractar cristales de proteína, también por darme la posibilidad de crear uno de los primeros cursos de bioinformática en la facultad de Ciencias, lo que ha sido una experiencia tremenda para mí. A las asistentes de investigación de nuestro grupo que siempre me dieron apoyo, Victoria Muñoz e Isabel Asela, su trabajo también fue importante para esta etapa.

A mis amigos y amigas del laboratorio, como no mencionar a las y los mandriles, son muchos años de ciencia, fútbol y celebraciones, por decirlo de modo educado. Seguro se me escaparan algunos, por eso no los mencionaré, pero sin duda creo que a la par de la tremenda formación que logré en este laboratorio está la satisfacción de haber conocido gente que estará toda la vida presente, más allá de la ciencia. Las anécdotas darían para escribir dos tesis como esta, por lo que prefiero dejarlas para la próxima vez que celebremos.

Como no mencionar en este espacio a aquellas personas que día a día trabajan junto a nosotros para que tengamos todo lo necesario en la universidad o en el laboratorio. A mi querida Sra. Nieves, por siempre escucharme y pese a no hacerme mucho caso, regalarme su amistad sincera y también las mejores sopaipillas que comimos en el lab. A don René, mi amigo, en que varias veces aburrido de pensar en proteínas me escapaba a su taller para conversar de la vida. A la querida Sra. Elizabeth, por soportar mis bromas, por escucharme y siempre estar preocupada por nosotros.

Esta tesis ha sido un camino largo, de profundo aprendizaje y formación, pero también de vivir experiencias y conocer personas que hicieron de ese camino algo más entretenido y valorable para el resto de la vida. ¡Gracias a todos y todas!

## **FINANCIAMIENTO**

- Beca de doctorado nacional CONICYT 21151101.
- Beca facultad de Ciencias Químicas, Universidad de Chile. Cierre de tesis doctoral.
- Proyecto FONDECY 1170701.

## PUBLICACIONES

### Publicaciones derivadas de esta tesis:

- Medina, E., **Villalobos, P.**, Hamilton, G. L., Komives, E. A., Sanabria, H., Ramírez-Sarmiento, C. A., & Babul, J. (2020). Intrinsically Disordered Regions of the DNA-Binding Domain of Human FoxP1 Facilitate Domain Swapping. *Journal of Molecular Biology.*, 432(19), 5411-5429.
- Medina, E., **Villalobos, P.**, Coñuecar, R., Ramírez-Sarmiento, C. A., & Babul, J. (2019). The protonation state of an evolutionarily conserved histidine modulates domain swapping stability of FoxP1. *Scientific Reports*, 9(1), 1-11.
- Medina, E., Córdova, C., **Villalobos, P.**, Reyes, J., Komives, E. A., Ramírez-Sarmiento, C. A., & Babul, J. (2016). Three-Dimensional domain swapping changes the Folding mechanism of the Forkhead domain of FoxP1. *Biophysical journal*, 110(11), 2349-2360.

### Otras publicaciones durante el desarrollo de esta tesis:

- Olavarria, K., Carnet, A., van Renselaar, J., Quakkelaar, C., Cabrera, R., da Silva, L. G., **Villalobos P.**, & Wahl, S. A. (2021). An NADH preferring acetoacetyl-CoA reductase is engaged in poly-3-hydroxybutyrate accumulation in *Escherichia coli*. *Journal of Biotechnology*, 325, 207-216.
- Zamora, R. A., Ramirez-Sarmiento, C. A., Castro-Fernandez, V., **Villalobos, P.**, Maturana, P., Herrera-Morandé, A., & Guixe, V. (2020). Tuning of Conformational Dynamics Through Evolution-Based Design Modulates the Catalytic Adaptability of an Extremophilic Kinase. *ACS Catalysis*, 10(19), 10847-10857.
- **Villalobos, P.**, Soto, F., Baez, M., & Babul, J. (2016). Regulatory network of the allosteric ATP inhibition of *E. coli* phosphofructokinase-2 studied by hybrid dimers. *Biochimie*, 128, 209-216.

## ÍNDICE DE CONTENIDOS

AGRADECIMIENTOS.....	iii
FINANCIAMIENTO .....	v
PUBLICACIONES.....	vi
ÍNDICE DE CONTENIDOS.....	vii
INDICE DE TABLAS .....	viii
ÍNDICE DE FIGURAS .....	ix
RESUMEN .....	xi
ABSTRACT .....	xiii
1. INTRODUCCIÓN.....	1
1.1. Características generales de la oligomerización de proteínas.....	1
1.2. El intercambio de segmentos como un mecanismo alternativo para generar oligómeros.....	3
1.3. Caracterización estructural, biofísica y evolutiva del intercambio de segmentos en la familia de factores de transcripción Fox.....	6
HIPOTESIS .....	14
OBJETIVO GENERAL.....	14
OBJETIVOS ESPECÍFICOS.....	15
2.MATERIALES Y MÉTODOS.....	16
2.1. Cepas y condiciones de cultivo.....	16
2.2 Mutagénesis sitio dirigida. ....	16
2.3. Sobreexpresión y purificación de las proteínas FoxP y sus mutantes.....	16
2.4. Cromatografía de exclusión molecular y cinéticas de disociación del dímero. ....	18
2.5. Análisis de las cinéticas de disociación. ....	18
2.6. Desplegamiento al equilibrio seguido por dicroísmo circular.....	20
2.7. Análisis del despliegamiento al equilibrio de las proteínas FoxP.....	21
2.8. Cristalización y difracción de cristales de FoxP4 .....	22
2.9. Análisis de identificación de residuos PCAIN. ....	23
2.10. Simulaciones moleculares para evaluar el rol de la protonación de H59 de FoxP1 .....	23
2.11. Simulaciones moleculares de los dímeros y monómeros FoxP. ....	24
2.12. Simulaciones de plegamiento de FoxP1.....	25
2.13. Filogenia y reconstrucción de secuencias ancestrales.....	26

3. RESULTADOS .....	27
3.1. Cinéticas de disociación de FoxP4 y FoxP1. ....	27
3.2. Determinación de la estructura de FoxP4.....	33
3.3. Estabilidad y mecanismo de plegamientos en la subfamilia FoxP. ....	39
3.4. Análisis de cambios en las redes de interacción en el núcleo hidrofóbico de las proteínas FoxP en el estado monomérico y dimérico.....	44
3.5. Caracterización de una mutante PCAIN. ....	48
3.6 Modulación del intercambio de segmentos por una histidina conservada en las familias FoxO, FoxM y FoxP. ....	51
3.7. Análisis de las mutantes derivadas de la comparación de secuencias FoxP.....	55
3.8. Dinámica conformacional de la familia FoxP. ....	60
3.9. Dinámica conformacional de los dímeros de FoxP2 y FoxP3. ....	69
3.10. Simulaciones moleculares de plegamiento del dímero de FoxP1.....	77
3.11. Análisis filogenético de la familia Fox humana. ....	85
3.12. Reconstrucción de Secuencias Ancestrales.....	89
4. DISCUSIÓN.....	96
4.1. La barrera entálpica de asociación determina las diferencias en la dimerización de FoxP4 y FoxP1.....	96
4.2. Un intermediario monomérico favorece la formación del intercambio por segmentos en la familia FoxP. ....	99
4.3 Los miembros de la subfamilia FoxP poseen una dinámica conformacional diferente.....	102
4.4. El intercambio de segmentos es modulado por residuos fuera de la región bisagra. ....	104
4.5. Las dos ramas filogenéticas de la subfamilia FoxP formaron dímeros con propiedades de oligomerización diferentes.....	107
5. CONCLUSIÓN .....	110
6. REFERENCIAS .....	112

## INDICE DE TABLAS

Tabla 1. Parámetros cinéticos y termodinámicos de las cinéticas de disociación del dímero de FoxP1 y FoxP4.....	31
Tabla 2. Parámetros de recolección de datos y refinamiento del monómero de FoxP4.....	37
Tabla 3. Parámetros termodinámicos de los experimentos de desplegamiento al equilibrio de las proteínas FoxP.....	43
Tabla 4. Parámetros cinéticos y termodinámicos de la disociación de la mutante PCAIN F7L.....	50

Tabla 5. Parámetros cinéticos y termodinámicos de disociación del dímero de las mutantes de FoxP1.....	59
Tabla 6. Abundancia de los <i>clusters</i> identificados en las simulaciones de los monómeros FoxP.....	63
Tabla 7. Correlaciones entre hélice H1 (R14) y carboxilo terminal de hélice H5.....	67
Tabla 8. Correlaciones entre hélice H1 (R14) y región entre H3-S2.....	67
Tabla 9. Abundancia de los clústeres de las simulaciones de los dímeros de FoxP2 y Foxp3. ....	72
Tabla 10 Correlaciones entre residuos inter-subunidad de las hélices H1. ....	76
Tabla 11. Correlaciones entre H1 (R14) y la región entre H3-S2 .....	76
Tabla 12. Parámetros obtenidos a partir de las cinéticas de disociación de AncP y FoxP1.....	95

## ÍNDICE DE FIGURAS

Figura 1. Esquema de la oligomerización por intercambio de segmentos .....	4
Figura 2. Secuencia y estructura del monómero y el dímero de la subfamilia FoxP .....	8
Figura 3. Cinética de disociación del dímero de FoxP4 a 25 °C. A.....	28
Figura 4. Barrera entálpica de la transición dímero monómero en FoxP1 y FoxP4. La constante de disociación ( $K_D$ ) y las constantes cinéticas descritas en la Tabla 1 fueron usadas para observar su dependencia con la temperatura.....	32
Figura 5. Cristales de proteína de FoxP4 en ausencia y presencia de ADN. ....	34
Figura 6. Estructura resuelta del monómero de FoxP4 y superposición estructural de los monómeros de la subfamilia FoxP .....	36
Figura 7. Desplegamiento al equilibrio de las proteínas FoxP .....	40
Figura 8. Redes de interacción inter-residuo en el núcleo hidrofóbico de las proteínas FoxP.....	46
Figura 9. Cinética de disociación del dímero de las mutantes PCAIN.....	49
Figura 10. Formación de puentes de hidrógeno entre H59 y N55 como consecuencia de la desprotonación de la histidina.....	53
Figura 11. Cinética de disociación del dímero de las mutantes de FoxP1. ....	58
Figura 12. Análisis de estabilidad conformacional de los monómeros FoxP. ....	62
Figura 13. Interfaz entre las hélices H1-H5 de los monómeros FoxP.....	64
Figura 14. Análisis de correlación de los movimientos atómicos de los monómeros FoxP.....	66
Figura 15. Análisis de las simulaciones moleculares de los dímeros de FoxP2 y FoxP3. ....	71
Figura 16. Matrices de correlación de las simulaciones moleculares de los dímeros de FoxP2 y FoxP3.....	75

Figura 17. Cambios de la fracción de los contactos nativos monoméricos (QM) y diméricos (QD) en las temperaturas de plegamiento TF1 y TF2. ....	79
Figura 18. Cambios en la capacidad calórica en las transiciones de plegamientos a diferentes concentraciones efectivas de proteína .....	80
Figura 19. Simulaciones moleculares del plegamiento del dímero de FoxP1.....	81
Figura 20. Cambios en las fracciones de las poblaciones de los ensambles conformacionales (N, I1, I2, U) en función de la temperatura y concentración de proteína .....	83
Figura 21. Alineamiento de secuencias de proteínas de la familia Fox de humanos. ....	87
Figura 22. Filogenia de la familia Fox humana. ....	88
Figura 23 Probabilidades de las secuencias de AncOP y AncP.....	92
Figura 24. Cinéticas de disociación del dímero de FoxP1 y el AncP.....	94
Figura 25. Origen y evolución del intercambio de segmentos en la familia Fox.....	111

## RESUMEN

Las proteínas Fox son una familia de factores de transcripción cuyo dominio de unión a ADN existe canónicamente como monómeros en solución. Sin embargo, las estructuras de FoxP2 y FoxP3, dos proteínas que son miembros de la subfamilia FoxP (FoxP1-4), las muestran como proteínas que dimerizan por intercambio de segmentos (DS). El DS es un mecanismo donde elementos idénticos de estructura secundaria forman dímeros estabilizándose por interacciones intermoleculares. Para determinar si los cambios de secuencia y estructura, observados en la subfamilia FoxP, pueden haber determinado la capacidad de dimerizar por este mecanismo, en el contexto de una familia monomérica, se realizó un análisis experimental y computacional de estas proteínas.

Experimentos de cinética de disociación del dímero de FoxP4, indican que tiene una afinidad intermedia de sus monómeros con una constante de disociación ( $K_D$ ) de  $135 \mu\text{M} \pm 8.0 \mu\text{M}$  a  $33^\circ\text{C}$ , respecto a los valores de  $K_D$  reportados para FoxP1 ( $\sim\mu\text{M}$ ), FoxP2 ( $\sim\text{mM}$ ) y a FoxP3 que es considerada como un dímero obligado. La comparación de FoxP4 con FoxP1 determinó que la barrera cinética entre FoxP1 y FoxP4, está principalmente explicada por una gran barrera entálpica de activación para la asociación de los monómeros, sugiriendo que las diferencias entre estas proteínas afectan los contactos entre los residuos que las conforman, definiendo por lo tanto su capacidad de dimerizar. Además, se obtuvo la primera estructura para FoxP4, que la muestra como un monómero a una resolución de  $2.2 \text{ \AA}$  en ausencia de ADN; la comparación estructural con el resto de los monómeros de FoxP1 y FoxP2 muestran que los monómeros tienen una alta similitud estructural. Para completar la caracterización biofísica, se realizaron experimentos de desplegamiento al equilibrio, seguidos por dicroísmo circular, a una concentración de proteína de  $20 \mu\text{M}$  para todas las proteínas FoxP. Los resultados mostraron que FoxP3 y FoxP1 tienen un mecanismo de desplegamiento de tres estados ( $N_2 \rightleftharpoons 2I \rightleftharpoons 2U$ ) con la acumulación de un intermediario monomérico (I) que retiene un  $\sim 80\%$  del contenido de estructura secundaria de la proteína. De acuerdo con la concentración de proteína usada, que en FoxP2 y FoxP4 favorece el monómero, ambas proteínas siguieron un mecanismo de desplegamiento de dos estados ( $N_2 \rightleftharpoons 2U$ ). Esos datos sugieren que las propiedades de dimerización están relacionadas con la presencia del intermediario monomérico.

Se evaluó la dinámica conformacional de la subfamilia FoxP con dinámicas moleculares convencionales utilizando las estructuras disponibles para los monómeros (FoxP1, FoxP2, FoxP4) y los dímeros (FoxP2 y FoxP3). Las simulaciones moleculares de los monómeros FoxP1 y FoxP4 y no las de FoxP2, mostraron movimientos concertados entre residuos del lazo entre H3-S2 y la hélice H1, lo que está de acuerdo a antecedentes de simulaciones de plegamiento que indican que esa vía de comunicación (H3-S2 y H1) es clave para la formación del estado de transición monomérico. Adicionalmente, durante las simulaciones moleculares de los dímeros de FoxP2 y FoxP3 se observa que solamente en FoxP3 se mantiene estable una interfaz secundaria entre las hélices H1 de ambas subunidades. De acuerdo con lo anterior, en FoxP1, que carece de esa interfaz, las simulaciones moleculares de plegamiento, utilizando modelos basados en estructura, muestran que la vía de disociación del dímero conecta al dímero nativo con dos ensamblajes conformacionales (I1 e I2) los que proveen en parte una explicación estructural de los intermediarios reportados utilizando FRET de molécula única.

Estudios filogenéticos de la familia Fox muestran que el intercambio de segmentos se origina con el ancestro de la subfamilia FoxP (AncP) y se relaciona a la mutación P39A en la región de la bisagra. Sin embargo, antes de que esta mutación se originara, ocurrieron otros cambios en la historia evolutiva de la familia que, secundariamente modulan el intercambio de segmentos: el estado de protonación de H59 modula el intercambio de segmentos según simulaciones moleculares de FoxP1 y el reemplazo de F7 (conservada en todas las proteínas FoxP) por leucina, aumenta la constante de disociación de FoxP1 24 veces. Finalmente, el árbol filogenético de la familia muestra a las proteínas FoxP divergiendo en dos ramas: una con el dímero obligado FoxP3 y otra que agrupa a FoxP1, FoxP2 y FoxP4, lo que está de acuerdo con sus propiedades de dimerización. Estos resultados sugieren que el intercambio de segmentos surgió en esta familia desestabilizando el monómero típico de la familia Fox y favoreciendo la formación de un intermediario monomérico que está estrechamente relacionado con la capacidad de dimerizar de estas proteínas, como también de su dinámica conformacional.

## ABSTRACT

The forkhead box (Fox) proteins are a widespread family of transcription factors whose DNA-binding domain exists canonically as a monomer in solution. Nevertheless, structures of FoxP2 and FoxP3, which are members of the P subfamily (FoxP1–4), showed dimeric structures via three-dimensional domain swapping (DS), a mechanism where the exchange of identical segments between subunits leads to intertwined dimers, stabilized by intermolecular interactions. To determine if small sequence and structure changes observed across this family may have differentially sculpted the dimerization ability, an experimental and computational analysis was performed.

Kinetic dissociation experiments of FoxP4 dimer indicated that this protein represents an intermediate affinity between monomers, with a dissociation constant ( $K_D$ ) of  $135 \mu\text{M} \pm 8.0 \mu\text{M}$  at  $33^\circ\text{C}$ , with respect to the values reported for FoxP1 ( $\sim\mu\text{M}$ ), FoxP2 ( $\sim\text{mM}$ ) and FoxP3 considered an obligated dimer. The comparison of FoxP4 with FoxP1 showed that the kinetic barriers between them are mainly explained due to the large enthalpic barrier of association of FoxP4 monomers, suggesting that the specific evolutionary differences affect residue contacts and therefore the dimerization ability of these proteins. The first crystal structure of monomeric FoxP4 at  $2.2 \text{ \AA}$  in absence of DNA was obtained, the structural comparison along with FoxP monomers showed high structural conservation with FoxP1 and FoxP2 monomeric structures. To complete the biophysical characterization, equilibrium unfolding experiments followed by circular dichroism were performed at the same protein concentration for all FoxP proteins ( $20 \mu\text{M}$ ). Results showed that FoxP3 and FoxP1 followed a three-state mechanism ( $N_2 \rightleftharpoons 2I \rightleftharpoons 2U$ ) with the accumulation of the monomeric intermediate (I) which retains  $\sim 80\%$  of the overall structural content ( $80\%$ ). According to the protein concentration used in the experiments, which favors the monomer of FoxP2 and FoxP4, both proteins followed a two-state folding mechanism ( $N \rightleftharpoons U$ ). All these data strongly suggest that the dimerization properties are related to the presence of a monomeric intermediary.

To gain insights about FoxP conformational dynamics, molecular dynamics simulations of the available FoxP monomeric (FoxP1, FoxP2, FoxP4) and dimeric (FoxP2, FoxP3) structures were performed. In line with previously reported experimental and folding

molecular dynamics studies, conventional simulations of FoxP monomers reveal correlated movement between residues located in a H3-S2 loop and H1 helix in FoxP1 and FoxP4, but not in FoxP2. Additionally, FoxP3 and FoxP2 dimer simulations show that the secondary interface between H1 helices of the FoxP3 dimer allows retaining the compact conformation of this protein. In agreement with the absence of the H1-H1 secondary interface, FoxP1 dimeric folding simulations using structure-based models shows that the dissociation pathway has locally disordered intermediaries (I1 and I2) which provide a structural description of experimental intermediaries observed by reported studies of single-molecule FRET. Altogether these results suggest that during evolution, conformational dynamics have a direct impact on native structure dynamics and trigger local unfolding events.

Phylogenetic studies of Fox family members show that the origin of domain swapping in the ancestor (AncP) is related to key mutation P39A in the hinge region of Fox proteins. However, before the origin of this mutation, other sequence changes emerged, as in the case of H59, which offers chemical control of the amino terminal of H3 helix depending on its protonation state, as molecular simulations dynamics simulations suggest. Also, replacement of a strictly conserved amino acid of FoxP proteins, F7, by leucine, observed in monomeric proteins, increases  $K_D$  over 24-fold. Interestingly, protonation of H59 or the F7L mutation in FoxP1 affects dimerization by increasing the kinetic barrier between the monomer and the transition state of association. Finally, the phylogenetic tree shows an exclusive branch emerging from AncP for the obligated dimer, FoxP3, and a second branch grouping FoxP1, FoxP4, and FoxP2, the FoxP members with reduced dimerization propensity. Altogether these results suggest that DS in the Fox family emerged by monomer destabilization allowing the emergence of a monomeric intermediary, which directly impacts the dimerization propensity of these proteins and that the evolutionary sequences changes along FoxP members modeled its conformational dynamics and dimerization propensities

## 1. INTRODUCCIÓN

### 1.1. Características generales de la oligomerización de proteínas.

La asociación de proteínas en complejos de dos o más cadenas polipeptídicas resulta en la formación de la estructura cuaternaria. La relevancia biológica del fenómeno de interacción proteína-proteína es evidente en virtud de la alta ocurrencia de este fenómeno y de sus consecuencias funcionales. Un estudio del año 2015 mostró que la clasificación de las estructuras resueltas por cristalografía de rayos-X en el banco de datos de proteínas (PDB) de acuerdo a su estado cuaternario ascendía a 7.972 monómeros, 9.206 homómeros (oligómeros compuestos de la misma cadena polipeptídica) y 2.677 heterómeros (oligómeros compuestos por cadenas polipeptídicas diferentes); es decir, 54% de las proteínas con estructura resuelta se asocian consigo mismas para formar homómeros (Marsh & Teichmann, 2015). A nivel celular, un estudio del proteoma de *Escherichia coli* en la base de datos SWISSPDB (actualmente UniProtKB/Swiss-Prot), en el cual no se consideraron los heterómeros, determinó que solo un 20% de las proteínas corresponden a monómeros (Goodsell & Olson, 2000).

La formación de estados oligoméricos permite la regulación de diversas funciones relevantes para la homeostasis celular, como la catálisis enzimática, la señalización celular, el transporte en canales iónicos y la regulación génica mediada por factores de transcripción (Marianayagam et al., 2004). Como consecuencia de ello, la oligomerización ha demostrado ser un factor clave en patologías que afectan diversos procesos celulares como el cáncer y enfermedades provocadas por agentes patógenos como los virus. Por ejemplo, las quinasas RAF regulan el crecimiento célula por medio del control de la vía de señalización ERK, y mutantes constitutivamente activas de esta proteína son responsables del desarrollo de tumores en humanos. La dimerización de RAF ha sido descrita como clave para el control de la cascada de señalización ERK, por lo cual su interfaz de dimerización ha sido propuesta como un blanco terapéutico para tratar el desarrollo de tumores (Rajakulendran et al., 2009) (Brummer & McInnes, 2020). Por otra parte, en el caso de la proteasa principal del virus SARS-CoV-2, se ha descrito que el dímero es la unidad activa y funcional que permite procesar las poliproteínas del virus mediante un corte proteolítico, liberando así las proteínas responsables de la replicación y ensamble del virus (Sacco et al., 2020). En efecto, esta proteasa ha sido objeto del diseño de drogas para controlar la enfermedad causada por el

virus SARS-CoV-2 (COVID19), por medio del diseño de fármacos que permitan impedir su dimerización y, en consecuencia, su función (Goyal & Goyal, 2020) (Sacco et al., 2020). Estos antecedentes muestran que la oligomerización de proteínas cumple un rol fundamental para controlar funciones biológicas relevantes, que constituye el blanco para el diseño de fármacos para el tratamiento de enfermedades en humanos.

Un aspecto clave de la oligomerización de proteínas, que ha sido objeto de extensa investigación a nivel biofísico y estructural, es conocer qué tan estable es la interacción entre las cadenas polipeptídicas de un oligómero y la esencialidad de esta interacción para la estabilidad de la estructura terciaria de cada subunidad. Es así como se han descrito proteínas en las que la asociación entre subunidades es fundamental para mantener su estructura nativa y función biológica, a las que se les conoce como oligómeros obligados o permanentes, mientras que los oligómeros no obligados no producen un deterioro drástico de la estructura nativa como consecuencia de la disociación. Estos últimos, por lo general, poseen constantes de disociación ( $K_D$ ) en el rango micromolar a milimolar (Nooren & Thornton, 2003).

A partir de lo anterior, surge la pregunta respecto a cuáles son los mecanismos evolutivos que establecen la formación de un oligómero a partir de proteínas monoméricas. Al respecto, se ha postulado que las proteínas monoméricas ancestrales acumularon un gran número de mutaciones en la superficie durante la evolución, generando una interfaz con otra cadena polipeptídica de forma tal que la energía libre liberada durante la oligomerización sea suficiente como para compensar el costo entrópico de eliminar los grados de libertad de rotación y traslación de las subunidades aisladas (Ma et al., 2003). De manera similar al rol del efecto hidrofóbico como un factor dominante de la formación del estado nativo, las mutaciones que se acumulan en la superficie proteica durante la evolución generan un parche hidrofóbico que favorece la formación de estados oligoméricos, en conjunto con la aparición de aminoácidos cargados que favorecen interacciones electrostáticas (Larsen et al., 1998). En este escenario, la estabilización de la interacción con otra cadena polipeptídica que contenga una superficie complementaria sigue principios fisicoquímicos que recapitulan el fenómeno de plegamiento de monómeros aislados (Janin et al., 1988).

La similitud entre los principios que dominan el proceso de plegado proteico y de la asociación proteína-proteína plantea, además, un mecanismo alternativo para la evolución

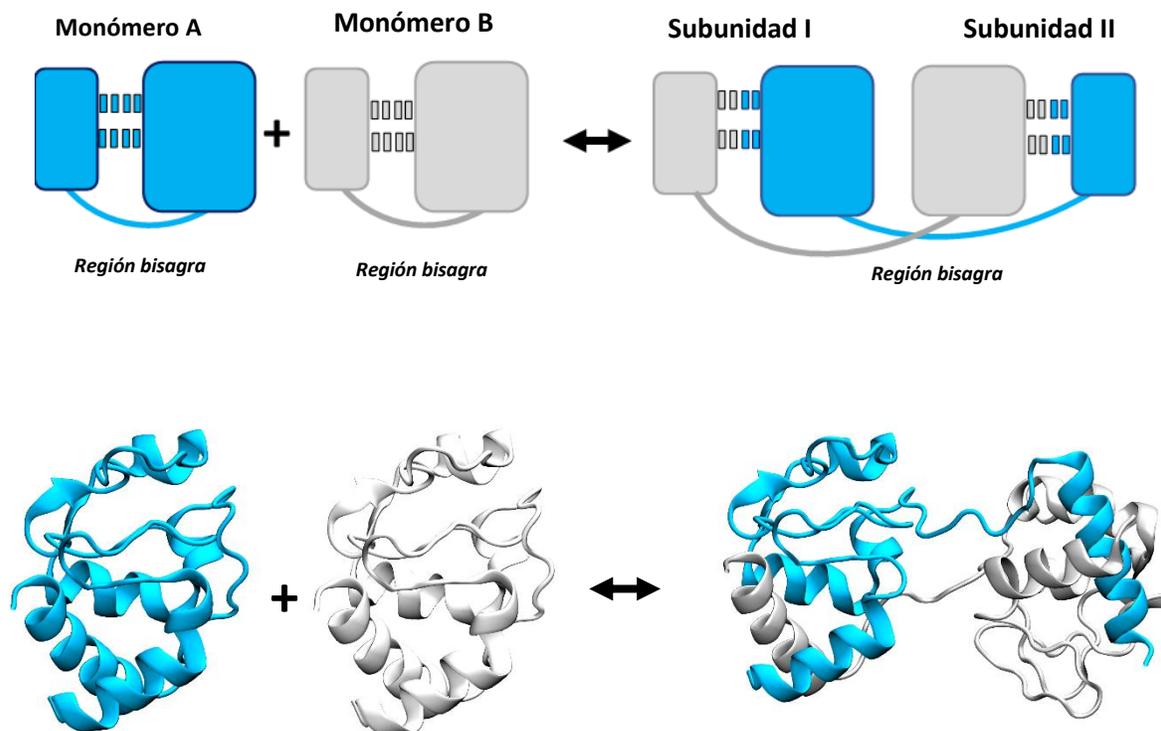
hacia formas oligoméricas, sin la necesidad de formar una nueva superficie de interacción: la posibilidad de establecer asociaciones entre dos o más cadenas polipeptídicas utilizando los mismos contactos que estabilizan al monómero, pero de forma intermolecular.

## **1.2. El intercambio de segmentos como un mecanismo alternativo para generar oligómeros**

El intercambio de segmentos (*domain swapping* en inglés) es un mecanismo alternativo al canónico de oligomerización de proteínas, en el cual se intercambian elementos de estructura secundaria o dominios entre dos o más cadenas polipeptídicas sin requerir de superficies especializadas de interacción. La ocurrencia de este mecanismo requiere que las interacciones que estabilizan al monómero aislado en solución se rompan y luego sean restituidas de manera intermolecular en el oligómero (Figura 1 A), de tal manera que la arquitectura tridimensional de un monómero aislado de una proteína se reconstituye en un contexto oligomérico, entrelazando la cadena polipeptídica (Bennet et al., 1995). Si bien los contactos del monómero aislado se mantienen en cada subunidad, pero de forma intermolecular en el dímero, existe una región en el oligómero que se diferencia respecto al monómero, generalmente denominada como región bisagra, la cual conecta a ambas subunidades en el dímero, (Yanshun & Eisenberg, 2002) (Figura 1 A).

Este mecanismo ha sido descrito para un gran número de proteínas de interés biológico, como es el caso de la proteasa principal del SARS-CoV (Zhong et al., 2009), citocromo *c* (Hirota et al., 2010), cistatina (Janowski et al., 2005), dominio DEP del señalosoma de Wnt (Gammons et al., 2016), entre otras. Gran parte de las pruebas experimentales que sostiene la presencia de oligómeros con intercambio de segmentos ha sido el reporte de las estructuras cristalográficas que presentan sus cadenas entrelazadas, lo cual ha llevado a cuestionar si es que este fenómeno puede ser observado en solución. En ese sentido, estudios de fluorescencia por transferencia de energía de resonancia de Förster (FRET) de molécula única de las repeticiones en tándem de dominios tipo inmunoglobulina (Ig), complementados por simulaciones moleculares, demostraron que estos dominios oligomerizan por intercambio de segmentos manteniendo regiones parcialmente desplegadas (Borgia et al., 2015). Además, experimentos de resonancia magnética nuclear en canecistatina-1 demostraron que los corrimientos químicos observados concuerdan con la

conformación por intercambio de segmentos resuelta en la estructura cristalográfica (Valadares et al., 2013).



**Figura 1. Esquema de la oligomerización por intercambio de segmentos.** Durante la asociación por intercambio de segmentos, dos monómeros (celestes y plomo) rompen parte de los contactos que los estabilizan y ahora esos contactos son reestablecidos, pero de manera intermolecular, con la subunidad adyacente en el dímero. Las líneas discontinuas representan los contactos intramoleculares (celestes o blanco) y los intermoleculares (celestes-blanco). La representación en cintas en la parte inferior de esta figura corresponde a las estructuras cristalográficas del monómero y del dímero con intercambio de segmentos de citocromo *c* de caballo, en donde se muestran a la izquierda dos monómeros (PDBid 1AKK) y a la derecha el dímero con intercambio de segmentos (PDBid 3WC8).

El requerimiento de la pérdida de interacciones que estabilizan al monómero aislado sugiere implícitamente que el desplegamiento de la proteína es necesario para alcanzar el estado oligomérico. En consistencia con esta aseveración, se ha demostrado experimentalmente que las proteínas con intercambio de segmentos necesitan tiempos de incubación muy extendidos para alcanzar el equilibrio monómero-oligómero en condiciones que favorecen el estado nativo. Uno de los casos más extremos le corresponde a cistatina, para la cual se ha estimado que requeriría cerca de 2.000 años para alcanzar el equilibrio monómero-dímero (Staniforth et al., 2001). En efecto, estudios en p13suc1, uno de los modelos mejor estudiados respecto al mecanismo de intercambio de segmentos, usando la especie monomérica aislada a una concentración de 0,5 mM, mostraron que se requirió de varios meses para alcanzar el equilibrio con su conformación dimerica con intercambio de segmentos. No obstante, el equilibrio entre especies se alcanzó rápidamente cuando se realizó el desplegamiento y replegamiento usando urea como agente caotrópico a la misma concentración de proteína (Rousseau et al., 2001). Asimismo, estudios de cinética de plegamiento de cianovirina-N confirmaron que la barrera de activación cinética es similar en magnitud a la barrera de desplegamiento, reforzando la idea de que el desplegamiento es necesario para este mecanismo (Liu et al., 2012). Finalmente, estudios *in silico* usando modelos tipo Gô con una función de energía que considera como interacciones atractivas entre residuos a nivel intra e intermolecular únicamente a los contactos nativos del monómero (*i. e.*, interacciones definidas entre residuos a partir de la estructura nativa de una proteína en función a un criterio de distancias), permitieron simular la formación del dímero con intercambio de segmentos de la proteína SH3 mediante del tránsito por el estado desplegado (Yang et al., 2004). De este modo, se ha establecido que el mecanismo de oligomerización por intercambio de segmentos requiere que el monómero alcance el estado desplegado y que la formación intermolecular de contactos propios del monómero es suficiente para la oligomerización.

El intercambio de segmentos es un caso interesante y atípico respecto a los mecanismos canónicos de oligomerización de proteínas, dado que no se requiere de la acumulación de mutaciones que permitan formar una nueva superficie de interacción. Dado que la región bisagra es el único segmento que difiere entre los estados monoméricos y oligoméricos de proteínas que se asocian por este mecanismo (Yanshun & Eisenberg, 2002),

tal que exista una conformación extensa o parcialmente desplegada que permita el intercambio de segmentos o dominios, se ha estudiado el efecto de la extensión y composición aminoacídica de esta región en diversos modelos proteicos de intercambio de segmentos. Respecto a la extensión, tanto la reducción del largo de la región bisagra en proteínas como p13suc1 (Rousseau et al., 2001), chkshs1 (Seeliger et al., 2002) y CD2 (Parker et al., 1998), como su extensión en la proteína CI2 (Chen et al., 1999), provoca una desestabilización del estado monomérico favoreciendo el dímero con intercambio de segmentos. Respecto a la composición aminoacídica, se cree que la presencia de residuos de prolina participa como un regulador del intercambio de segmentos (Bergdoll et al., 1997). En el caso de la proteína p13suc1, la substitución de Pro90 desplaza el equilibrio hacia el monómero, mientras que Pro92 lo hace hacia el dímero (Rousseau et al., 2001). No obstante, también se ha comprobado el rol de residuos distantes de la bisagra en el equilibrio dímero-monómero en casos como p13suc1, en la cual mutaciones por alanina en un sitio de unión a fosfato y que está alejado de la bisagra desplazó el equilibrio hacia monómero, conduciendo a cambios en  $K_D$  de dos órdenes de magnitud mayores (Rousseau et al., 2001). Basándose en que la estructura del monómero y el dímero difieren estructuralmente solo en la región bisagra, se postuló que los cambios observados en el equilibrio dímero-monómero provocado por mutaciones localizadas en diferentes regiones alejadas de la bisagra estarían comunicando ambas regiones, perturbándola y facilitando el intercambio de segmentos (Schymkowitz et al., 2001).

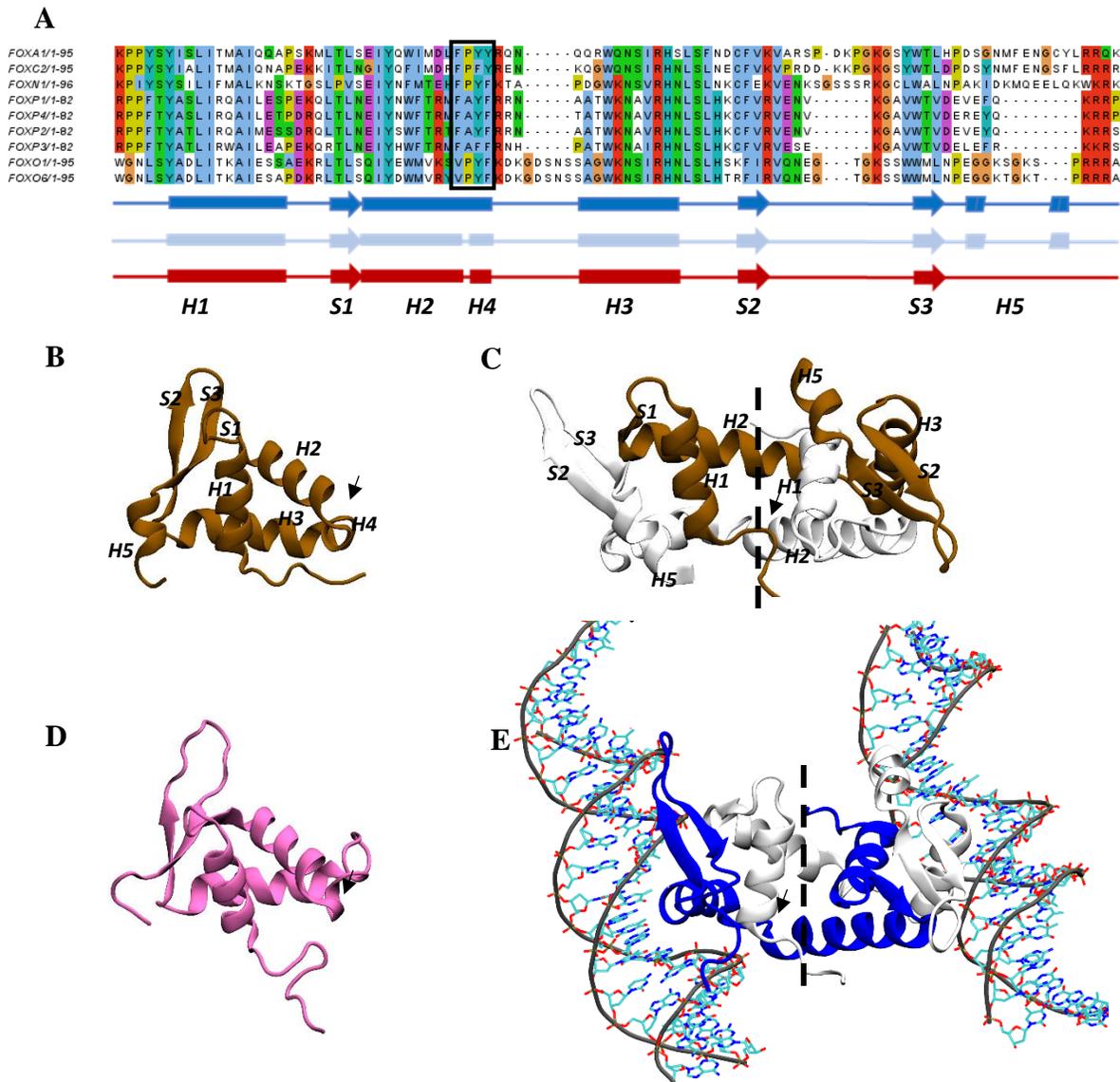
### **1.3. Caracterización estructural, biofísica y evolutiva del intercambio de segmentos en la familia de factores de transcripción Fox**

Si bien se han podido corroborar experimental y computacionalmente las reglas fundamentales del mecanismo de oligomerización por intercambio de segmentos en un escenario de desplegamiento y replegamiento de proteínas en altas concentraciones, su relevancia biológica ha sido cuestionada en virtud de los requerimientos atípicos para facilitar este mecanismo de oligomerización *in vitro*. No obstante, se ha descrito una familia de factores de transcripción que escapa de los principios previamente descritos para las proteínas que oligomerizan formando un dímero por intercambio de segmentos.

Los miembros de la familia de factores de transcripción *forkhead box* se encuentran asociados a diversos procesos celulares que involucran la regulación del metabolismo,

apoptosis, proliferación, diferenciación, y progresión del ciclo celular, entre otros (Jackson et al., 2010). Poseen un dominio de unión a ADN (denominado Fox en adelante, Figura 2) de aproximadamente 100 aminoácidos que está altamente conservado en todos sus miembros. Los factores de transcripción Fox que han sido estudiados a nivel funcional, estructural y biofísico de manera más extensa son aquellos encontrados en humanos, y la comparación de las secuencias de esos dominios en la familia permite su clasificación en 19 subfamilias, (FoxA a FoxS) (Jackson et al., 2010). Respecto a su estructura, el dominio de unión a ADN es una variante del plegamiento hélice-lazo-hélice presente en histonas y contiene tres hélices  $\alpha$  (H1, H2 y H3), tres hebras  $\beta$  (S1, S2 y S3) y dos lazos (W1 y W2), elementos que se encuentran conectados de la siguiente forma: H1-S1-H2-H3-S2-W1-S3-W2. De estos elementos, la hélice H3 es la encargada de contactar al surco mayor del ADN (Clark et al., 1993), mientras que los lazos W1 y W2 han sido descritos como importantes para el reconocimiento del ADN y para la interacción con otras proteínas (Clark et al., 1993). Varias de las estructuras de los miembros de la familia Fox muestran a estos factores de transcripción como monómeros, como es el caso de FoxA2 (PDBid: 5X07) (Li et al., 2017), FoxO3 (PDBid: 2UZK) (Tsai et al., 2007) y FoxN3 (PDBid: 6NCE) (Rogers et al., 2019), entre otras. Solo la subfamilia FoxP muestra pruebas clara, tanto a nivel estructural como en solución, de que las proteínas de esta subfamilia pueden oligomerizar por intercambio de segmentos (Stroud et al., 2006) (Bandukwala et al., 2011).

La primera estructura de la subfamilia FoxP, cuyos miembros son FoxP1-4, fue resuelta para FoxP2 en presencia de ADN, conteniendo tanto la especie monomérica como dimérica en la misma unidad asimétrica del cristal, y mostrando que el dímero corresponde a un oligómero con intercambio de segmentos, en el cual los sitios de unión a ADN se encuentran en extremos opuestos de la proteína (Stroud et al., 2006). Las estructuras cristalográficas resueltas posteriormente para FoxP3 corresponden a un dímero con intercambio de segmentos interactuando con NFAT (Bandukwala et al., 2011) y a un dímero aislado (Chen et al., 2015), en ambos casos uniendo dos moléculas de ADN en extremos opuestos.



**Figura 2. Secuencia y estructura del monómero y el dímero de la subfamilia FoxP.** En A se muestra un alineamiento múltiple de las secuencias de la subfamilia FoxP con algunos miembros de la familia Fox. La secuencia destacada con un rectángulo negro corresponde a la región bisagra que permite el intercambio de segmentos en el dímero. Bajo el alineamiento de secuencias, se representa la topología de elementos de estructura secundaria, altamente conservada en el plegamiento de la familia. Solo se observan pequeños cambios en la región del carboxilo terminal y la región bisagra de la subfamilia FoxP al oligomerizar por intercambio de segmentos. En B y C, se muestran las estructuras resueltas del monómero y dímero con intercambio de segmentos de FoxP2 (PDBid 2A07), respectivamente. En D y E, se muestra el monómero de FoxA2 (PDBid: 5X07) y el dímero con intercambio de segmentos de FoxP3 (PDBid: 3QRF). En el caso de FoxP3, también se presentan en representación de líneas dos moléculas de ADN con las cuales interactúa simultáneamente el dímero con intercambio de segmentos.

En el caso del dímero de las proteínas FoxP, los elementos que se intercambian con la subunidad adyacente en comparación al empaque del monómero aislado son las hélices H2 y H4, las hebras S2 y S3 y los lazos W1 y W2 (Figura 2 B y 2 C). Del análisis de las estructuras y secuencias de FoxP, se determinó que el intercambio de segmentos en esta subfamilia es producto de la sustitución de un residuo de prolina por alanina en el motivo FPYF, altamente conservado entre los distintos miembros de Fox, lo cual permite la elongación de la hélice H2 del dímero de FoxP en dos vueltas (Figura 2 A), facilitando así el intercambio de segmentos (Stroud et al., 2006). En base a esos antecedentes y al posicionamiento de los sitios de unión a ADN en el oligómero (Figura 2 E), se propuso que el intercambio de segmentos facilitaría la unión de la proteína a sitios distantes en una misma doble hebra de ADN o en dos moléculas de ADN diferentes, de tal manera que el oligómero contribuiría al establecimiento de interacciones cromosómicas de largo alcance (Stroud et al., 2006).

Esta proposición fue estructural y funcionalmente validada con la resolución de la estructura de FoxP3 en presencia de ADN (Figura 2 E), en donde se muestra a esta proteína contactando a dos moléculas de DNA de forma anti paralela (Bandukwala et al., 2011) (Chen et al., 2015). En esa línea, experimentos de interacción cromosomal de largo alcance mediante la técnica 4C-seq, que permite secuenciar regiones del ADN que se encuentran próximas en el espacio como consecuencia de la interacción con factores de transcripción, logró comprobar que FoxP3 facilita la formación de contactos intra e inter cromosomales (Chen et al., 2015). Complementariamente, la relevancia biológica del mecanismo de intercambio de segmentos ha sido mayormente reflejada en el hallazgo de mutaciones en FoxP2 y FoxP3 relacionadas a afecciones al desarrollo del lenguaje y a un síndrome congénito de inmunodeficiencia (IPEX), respectivamente, las cuales causan un impedimento de la formación del dímero entrelazado (Bandukwala et al., 2011). Adicionalmente, experimentos *in vivo* con una mutante de FoxP3 que favorece la especie monomérica de la proteína, mostraron que la modificación afecta la transcripción de genes cuya expresión es regulada por la acción de FoxP3 (Bandukwala et al., 2011).

La relevancia biológica del dímero de FoxP3 contrasta fuertemente con el mecanismo molecular propuesto para la oligomerización por intercambio de segmentos, debido a que

esto último ha sido descrito extensamente como un fenómeno de cinética lenta en ausencia de agentes caotrópicos. Esto propone la necesidad de una revisión exhaustiva del mecanismo de intercambio de segmentos, para poder conciliar los aspectos biofísicos y funcionales de este proceso en la subfamilia FoxP. En esa línea, experimentos de dispersión de luz multiangular sobre FoxP2 demostraron que este dímero se forma en solución en ausencia de efectores o ligandos, y para FoxP2 el equilibrio de este proceso se alcanza en 24 h a 25 °C y en 10 h a 37 °C (Stroud et al., 2006). Por otra parte, las pequeñas variaciones de secuencia en la subfamilia FoxP en regiones alejadas de la bisagra son relevantes para la estabilidad del estado oligomérico. Por ejemplo, mientras que FoxP2 tiene una  $K_D$  de dimerización de ~2,4 mM (Perumal et al., 2015), ensayos de cromatografía de exclusión molecular sobre FoxP1 indican consistentemente que dimeriza con una  $K_D$  de orden micromolar luego de solo minutos de incubación y en ausencia de agentes caotrópicos (Medina et al., 2016). En contraste, FoxP3 parece ser un dímero obligado, ya que no es posible determinar la presencia del monómero aislado en solución ensayando concentraciones de proteína del orden micromolar (Bandukwala et al., 2011). Respecto a FoxP4, cabe señalar que no existen estudios biofísicos que describan sus propiedades de oligomerización y se desconoce cuál es su estado oligomérico.

Estos antecedentes sugieren fuertemente que el proceso de oligomerización de FoxP es energéticamente favorable y mecanísticamente disímil del propuesto para modelos canónicos de intercambio de segmentos, como p13suc1, que requieren de perturbaciones como cambios de pH, temperatura u otros agentes caotrópicos (Rousseau et al., 2001). La reciente exploración del plegamiento y oligomerización al equilibrio de FoxP1 (Medina et al., 2016) demostró que el proceso de intercambio de segmentos posee un cambio en energía libre de disociación altamente desfavorable ( $\Delta G_D \sim 7$  kcal/mol). Además, este trabajo determinó que, a diferencia de los modelos clásicos de intercambio de segmentos cuyos dímeros siguen un mecanismo de plegamiento de dos estados  $N_2 \rightleftharpoons 2U$  (Staniforth et al., 2001), FoxP1 sigue un mecanismo de plegamiento de tres estados ( $N_2 \rightleftharpoons 2I \rightleftharpoons 2U$ ), en donde la disociación del dímero ocurre previo al desplegamiento de la proteína completa y que es acompañado por cambios discretos en el contenido de estructura secundaria y en el radio hidrodinámico. Adicionalmente, experimentos de intercambio de hidrógeno/deuterio medidos por espectroscopia de masas sobre el monómero nativo y el intermediario

monomérico, cuya población alcanza un máximo a concentraciones bajas (~1,8 M) del agente caotrópico cloruro de guanidinio, comprobaron que la estructura del intermediario es muy semejante a la del monómero nativo FoxP1, sugiriendo fuertemente que el proceso de oligomerización de FoxP1 ocurre a partir de monómeros extensamente estructurados (Medina et al., 2016).

La noción del intercambio de segmentos como un mecanismo de evolución de oligómeros en familias de proteínas, ha sido estudiada en varias familias de proteínas como es el caso de BsRnasa A y las  $\gamma$ -cristalinas (Bennet et al., 1995). Desde un punto de vista evolutivo, la presencia de monómeros y dímeros en la subfamilia FoxP y también considerando que la familia Fox está compuesta esencialmente monómeros, es claro que el intercambio de segmentos se originó a partir de un ancestro monomérico en esta familia. Por el contrario, un ejemplo de especial interés es el caso de las histonas, que forman la arquitectura básica de las proteínas que componen el nucleosoma en eucariontes y cuya estructura funcional corresponde exclusivamente a dímeros que oligomerizan por intercambio de segmentos, sin existir la evidencia experimental de histonas que se pliegan como monómeros. Interesantemente, los factores de transcripción Fox e histonas tienen una considerable semejanza (Clark et al., 1993) debido a que ambos son variantes del motivo estructural hélice-lazo-hélice (Lalmansingh et al., 2012),

Para investigar el origen del intercambio de segmentos en las histonas se realizó un estudio una búsqueda de secuencias de homólogos distantes de histonas identificó el dominio C-terminal de las ATPasas AAA+, el cual se superpone estructuralmente con una subunidad del dímero de las histonas, pero que cuenta con una inserción de glicina, treonina y prolina (motivo GTP) en la región correspondiente a la  $\alpha$ -hélice que se intercambia entre ambas subunidades en histonas. Estos residuos son reconocidos por impedir la formación de estructura helicoidal y, consistentemente, permitían el empacamiento de una cadena polipeptídica aislada en el caso del C-terminal de las ATPasas AAA+ (Alva et al., 2007). La inserción de ese motivo en la bisagra de una histona dimérica de arquea (HMfB) resultó en un monómero soluble y estable (Hadjithomas & Moudrianakis, 2011), sugiriendo fuertemente que, durante la evolución de estas proteínas, el surgimiento del intercambio de segmentos en un oligómero a partir de un ancestro monomérico mediante cambios locales de

secuencia permitió la aparición de una nueva familia de proteínas de trascendencia biológica como lo son las histonas. Ambos casos, histonas y factores de transcripción FoxP, son ejemplos de cómo por medio de la adopción de mínimos cambios de sus secuencias durante la evolución se estabilizó un dímero con intercambio de segmentos a partir de proteínas monoméricas.

En contraste con lo ocurrido para histonas, diversos antecedentes experimentales sugieren que los cambios evolutivos ocurridos en FoxP para esculpir la oligomerización mediante el mecanismo de intercambio de segmentos son más complejos. Pese a que se ha definido que la bisagra juega un rol trascendental para la formación de especies oligoméricas vía intercambio de segmentos (Stroud et al., 2006) (Medina et al., 2016), un intento por recapitular el cambio de prolina por alanina que ocurrió en el motivo conservado FPYF de las proteínas FOX sobre la proteína monomérica FOXK1 no permitió la formación de una especie oligomérica incluso ensayando concentraciones de proteína de 2 mM (Chu et al., 2011). Se trató lo inverso en FoxP3, reemplazando la alanina por prolina en la región bisagra (reestableciendo el motivo FPYF) y adicionando otras dos mutaciones en base a la interfaz del intercambio, sin embargo, no fue posible obtener una población completamente monomérica (Bandukwala et al., 2011) como en FoxP2 (Stroud et al., 2006), lo que augura la existencia de otros determinantes estructurales que favorezcan el intercambio de segmentos en la evolución de la familia Fox. Al mismo tiempo, experimentos de Resonancia magnética Nuclear (RMN) en FoxP1 mostraron que la región de la aleta W1 y C-terminal son flexibles en escala de ps-ns, mientras que la bisagra tiene un cambio conformacional más lento, en la escala de tiempo de  $\mu$ s-ms (Chu et al., 2011), sugiriendo que estas diferencias de flexibilidad en sitios distales de la región de entrelazado pueden ser importantes para la oligomerización en la subfamilia FOXP, tal como se observó en p13-suc1 (Schymkowitz et al., 2001).

En base a esos antecedentes, es posible proponer que, durante la evolución de la familia Fox, el intercambio de segmentos surgió producto de la mutación de alanina por prolina en la región que se conoce como región bisagra y de cambios de secuencia entre las secuencias FoxP que permitieron modelar las propiedades de dimerización de cada una de estas proteínas. Contrario a lo observado en los mecanismos canónicos de intercambio de

segmentos, donde el desplegamiento total es necesario, las proteínas FoxP probablemente se ven afectadas por fenómenos de desplegamiento local o de alta heterogeneidad conformacional que les permite asociarse a otra cadena.

En ese escenario, en este trabajo de tesis se propuso un estudio evolutivo para comprender el surgimiento de la oligomerización por intercambio de segmentos en la subfamilia FoxP. Primero, para comprender las diferencias en el mecanismo de oligomerización de la subfamilia FoxP, se realizó una completa caracterización biofísica estructural de FoxP4, el único miembro de esta subfamilia no caracterizado hasta la fecha. Esto consistió en la determinación de la cinética y termodinámica de disociación de FoxP4 en comparación al modelo de estudio FoxP1, y experimentos de cristalografía de rayos X que permitieron determinar exitosamente la estructura tridimensional de FoxP4. Esto fue acompañado de estudios de desplegamiento al equilibrio de los cuatro miembros de la subfamilia FoxP, para determinar la relación entre oligomerización y plegamiento en estas proteínas, y la elaboración de un *script* de análisis de estructura de proteínas que permite comprender las diferencias en las redes de interacción atómica en la estructura de las proteínas de esta familia. Finalmente, se realizó un análisis filogenético de la familia, que permitieron reconstruir la secuencia del último ancestro en común de la familia P con el resto de las proteínas monoméricas (AncOP) y el último ancestro en común de la subfamilia FoxP. En base al análisis estructural y las diferencias en secuencia de los miembros de la subfamilia FoxP, y se realizaron mutaciones para confirmar los resultados del análisis estructural y evolutivo, usando como modelo de estudio a FoxP1. Este trabajo fue complementado por simulaciones computacionales de dinámica molecular usando aproximaciones clásicas y de grano grueso para explicar las transiciones conformacionales a mayor escala.

## **HIPOTESIS.**

Durante la evolución de la familia Fox, la estabilización energética de un dímero con intercambio de segmentos a partir de formas ancestrales de menor afinidad en la subfamilia ha ocurrido mediante cambios tanto a nivel de la región bisagra, que permite el intercambio, como también en aminoácidos distantes de esta región. Estos cambios han esculpido la dinámica conformacional de las proteínas de la familia FoxP y con ello facilitado la desestabilización del monómero y concomitantemente la formación del dímero con intercambio de segmentos.

## **OBJETIVO GENERAL.**

Describir los cambios evolutivos necesarios para permitir la oligomerización mediante intercambio de segmentos a partir del ancestro común de la subfamilia FoxP y para explicar los cambios en la constante de disociación del dímero entre los miembros actuales.

## **OBJETIVOS ESPECÍFICOS.**

1. Determinar la filogenia de la familia FOX y reconstruir las secuencias y estructuras ancestrales de dos nodos del árbol, el último ancestro en común entre las proteínas FoxP (AncP), utilizando los métodos de máxima verosimilitud y modelamiento por homología.
2. Describir el proceso de dimerización de FoxP4 y el ancestro AncP utilizando cromatografía de exclusión molecular.
3. Caracterizar biofísicamente y en diversos niveles de resolución a FoxP4, FoxP3 y AncP, utilizando dicroísmo circular, cristalografía de rayos X.
4. Recapitular los cambios evolutivos necesarios para modificar la afinidad de la interacción proteína-proteína utilizando a FoxP1, el factor de transcripción caracterizado en nuestro laboratorio, como modelo de estudio.
5. Conciliar los resultados del análisis evolutivo y biofísico del proceso de dimerización y los cambios de flexibilidad conformacional local de las proteínas ancestrales y actuales de la subfamilia FOXP mediante estrategias *in silico* usando modelos de simulación de grano grueso y simulaciones moleculares convencionales.

## 2.MATERIALES Y MÉTODOS

### 2.1. Cepas y condiciones de cultivo.

Las cepas utilizadas durante el desarrollo de esta tesis se obtuvieron del Laboratorio de Bioquímica y Biología Molecular, Facultad de Ciencias de la Universidad de Chile. La cepa de *E. coli* DH5 $\alpha$  fue utilizada para la transformación y extracción de plásmidos. La cepa de *E. coli* y BL21(DE3)-C41 fue utilizada para inducir la sobreexpresión de la proteína debido a que su genoma posee una copia del gen codificante de la ARN polimerasa del fago T7, el que es controlado por el operón *lac* y cuya expresión es inducible por IPTG (isopropil- $\beta$ -D-tiogalactopiranosido). Esta cepa contiene el plásmido pET-TEV-sFoxP que codifica para la proteína de interés (FoxP1, FoxP4, FoxP3 o mutantes de FoxP1) y posee resistencia a kanamicina. Previamente a la secuencia FoxP se ubican tres secuencias: una etiqueta de polihistidina para facilitar la purificación por cromatografía de afinidad por níquel, un sitio de reconocimiento de la proteasa TEV (*Tobacco Etch Virus*), y el péptido S de la RNAsa A, que favorece la solubilidad sin alterar la estructura de la proteína (Kim & Raines, 1993).

### 2.2 Mutagénesis sitio dirigida.

Los partidores fueron sintetizados por la compañía IDT (*Integrated DNA Technologies*) y diseñados de acuerdo con las instrucciones del protocolo para mutagénesis sitio dirigida Gene Tailor™. El plásmido pET-TEV-sFoxP fue utilizado como molde para realizar la reacción en cadena de la polimerasa (PCR) usando la enzima Pfu Ultra II (Agilent Technologies). La enzima *DpnI* fue utilizada para digerir el producto de amplificación durante 1 h, y una alícuota de esa digestión fue utilizada para la transformación de *E. coli* DH5 $\alpha$ . Los clones fueron seleccionados en medio agar-LB (Luria-Bertani), aislados utilizando el protocolo comercial E.Z.N.A Plasmid Kit (Omega Bio-tek, Georgia, USA) y secuenciados (Macrogen Inc, Rockville, USA) para corroborar los cambios en la secuencia.

### 2.3. Sobreexpresión y purificación de las proteínas FoxP y sus mutantes.

Para todas las proteínas se utilizó el protocolo reportado de sobreexpresión y purificación para FoxP1 (Medina et al., 2016) y en el caso de haber incluido variaciones en éste, serán debidamente indicadas. Las células BL21(DE3)-C41 que contienen los respectivos plásmidos sFoxP fueron crecidas a agitación constante durante 16 h a 37 °C en 50 mL de medio LB suplementado con kanamicina, luego 40 mL fueron utilizados para

inocular 2 L de medio Terrific (12 g/L, extracto de levadura 24 g/L y glicerol al 0,4 % v/v) suplementado con el antibiótico. Ese cultivo fue sometido a agitación constante a 37 °C hasta alcanzar una densidad óptica de ~0,9, la que fue usada como referencia para inducir la sobreexpresión por la adición de IPTG a una concentración final de 0,5 mM e incubando el cultivo a 25 °C por 16 h (solo en el caso de FoxP4 la sobreexpresión fue inducida a 20 °C). El cultivo fue centrifugado en un rotor F12-6x500lex (Thermo Scientific) a 6.000 rpm a 4 °C por 30 min para obtener el sedimento bacteriano. El sedimento fue resuspendido en 100 mL en solución A (HEPES 20mM pH 7,8, NaCl 500 mM,  $\beta$ -mercaptoetanol 2 mM, imidazol 30 mM) y fue sometido a lisis celular por ultrasonido en un sonicador digital Sonifier 450 (Branson) en 12 pulsos con amplitud 40% durante 20 s con intervalos de pausa de 60 s. El contenido celular luego se centrifugó por 20 min a 18.000 rpm en un rotor F21-8x50y (Thermo Scientific), y el líquido sobrenadante fue colectado, siendo cargado en una columna HisTrap HP de 5 mL (GE Healthcare Bio-Scienies) equilibrada con solución A. La columna con la proteína cargada fue lavada con 50 mL de solución A y luego eluída con un gradiente de 50 mL entre la solución A y B (HEPES 20 mM, pH 7,8, NaCl 500 mM,  $\beta$ -mercaptoetanol 2 mM, imidazol 500 mM) en un equipo Äkta (General Electric) siguiendo la absorbancia a 280 nm. La proteína fue colectada y suplementada a una concentración final de 0,5 mM de EDTA.

Finalmente, la proteasa TEV (Laboratorio de Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad de Chile) fue utilizada para escindir la etiqueta de poli-histidina, realizando una incubación durante 16 h a 4°C usando una proporción 1:20 (TEV: proteína total). Luego, el contenido fue centrifugado a 4 °C por 10 min a 18.000 rpm en usando el rotor F12-6x500lex (Thermo Scientific), y el contenido fue sometido a diálisis por 4 h a 4 °C y posteriormente centrifugado en las condiciones indicadas anteriormente. La purificación de la proteína fue realizada en la columna Histrap-HP de 5 mL colectando el frente de carga. Las proteínas sin la etiqueta de poli-histidina fueron concentradas a ~0,6 mg/mL, desplegadas en solución de almacenaje (NaH<sub>2</sub>PO<sub>4</sub> 20 mM, pH 7.0, NaCl 100 mM y cloruro de guanidinio (GndCl) 4M) y guardadas a 4 °C.

## 2.4. Cromatografía de exclusión molecular y cinéticas de disociación del dímero.

Los experimentos de exclusión molecular fueron realizados en un equipo HPLC Waters 1525 equipado con un detector Waters 2487 (Waters Corp) al que se conectó a una columna Superdex 75 10/300 GL (GE Healthcare Bio-Sciences). La columna fue equilibrada con solución C ( $\text{Na}_2\text{HPO}_4$  20 mM, pH 7,  $\text{Na}_2\text{SO}_4$  100 mM,  $\beta$ -mercapotetanol 2 mM). La columna fue calibrada con un estándar Bio-Rad, el que permitió determinar los volúmenes de retención respecto al estándar que contiene muestras de proteínas con peso molecular conocido: vitamina de  $\text{B}_{12}$  (masa molecular 1,35 kDa y radio hidrodinámico 8,5 Å), mioglobina de caballo (masa molecular 17 kDa y radio hidrodinámico 19 Å), y ovoalbúmina de pollo (masa molecular 44 kDa y radio hidrodinámico 30,5 Å).

Los experimentos de cinéticas de disociación del dímero fueron realizados concentrando la proteína de interés a una concentración de 4-7 mg/mL y luego cargando esa proteína en la columna para separar el dímero del monómero. La elución fue seguida a dos longitudes de onda (220 y 280 nm) para coleccionar la fracción correspondiente al dímero, el que fue preparado a una concentración de 3  $\mu\text{M}$  para FoxP1 y 39  $\mu\text{M}$  para FoxP4. Las proteínas fueron incubadas a 20, 25, 30 y 33 °C en el caso de FoxP4 y a 33, 37, 40 y 45 °C en el caso de FoxP1, colectando alícuotas a diferentes tiempos. En el caso de FoxP4 las muestras fueron tomadas en un rango entre 0 y 28 h y en FoxP1 0 y 3 h. Cada alícuota fue posteriormente cargada en la columna Superdex 75 10/300 GL y los cromatogramas obtenidos fueron ajustados a una función de dobles gaussianas que permitió cuantificar la fracción de dímero y monómero en cada tiempo del ensayo. Las fracciones del dímero fueron graficadas en función el tiempo y ajustadas a una función de decaimiento exponencial simple (ecuación 1), como se explica la siguiente sección.

## 2.5. Análisis de las cinéticas de disociación.

Los datos de la variación del porcentaje de dímero en función del tiempo fueron ajustados a una función de decaimiento exponencial simple, la que se describe en la siguiente ecuación:

$$Y = (Y_0 - \text{Plateau}) \cdot e^{(-kt)} + \text{Plateau} \quad (1)$$

Donde  $Y_0$  corresponde al % de dímero en el tiempo cero, la meseta corresponde al valor de la función a tiempo infinito (cuando la reacción de disociación alcanza el equilibrio), y  $k$  es la constante cinética de disociación ( $k_d$ ). Asumiendo un modelo simple de disociación:



Donde D corresponde al dímero y M al monómero, la constante de disociación al equilibrio de esa reacción,  $K_D$ , se define como:

$$K_D = \frac{[M]^2}{[D]} \quad (3)$$

Por otro lado, la concentración de dímero de cada medición se puede relacionar a su respectiva fracción como sigue:

$$[D] = \frac{f_D \cdot [P_T]}{2} \quad (4)$$

Donde  $f_D$  corresponde a la fracción de dímero, y  $P_T$  es la proteína total. De manera alternativa a lo mostrado en la ecuación (3), la constante de disociación ( $K_D$ ), puede ser definida utilizando  $f_D$  y la concentración de proteína total ( $P_T$ ) como sigue:

$$K_D = \frac{2((1-f_D)^2 \cdot ([P_T]))}{f_D} \quad (5)$$

Las constantes cinéticas,  $k_d$  y  $k_a$ , corresponden a la constante de velocidad de disociación y a la constante de asociación de los monómeros, respectivamente. Ambas se pueden relacionar con la constante de disociación,  $K_D$ , mediante la siguiente ecuación:

$$K_D = \frac{k_d}{k_a} \quad (6)$$

De esta forma, los valores obtenidos de los experimentos de disociación ( $K_D$  y  $k_d$ ) permiten extrapolar el valor de la constante cinética de asociación,  $k_a$ . La descripción de la constante de disociación se relaciona con la diferencia de energía libre ( $\Delta G$ ) de la reacción de la siguiente forma:

$$\Delta G = RT \ln K_D \quad (7)$$

Donde  $R$  y  $T$  son la constante de los gases ( $8,314 \text{ J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$ ) y la temperatura (en grados Kelvin), respectivamente.

La evaluación de la constante de disociación  $K_D$  en función de la temperatura permite obtener la contribución entálpica y entrópica del proceso de disociación al equilibrio, lo que se describe en ecuación de Van't Hoff:

$$\ln K_D = \frac{-\Delta H}{RT} + \frac{\Delta S}{R} \quad (8)$$

En donde  $\Delta H$  y  $\Delta S$  corresponden a las diferencias de entalpía y entropía, respectivamente.

Para el caso de la contribución de los parámetros antes mencionados en la cinética de disociación, las constantes cinéticas fueron analizadas en términos de la ecuación de Eyring (Eyring, 1935) para determinar el cambio en energía libre de que separa a monómeros y dímeros con el estado de transición,  $\Delta G_a^\ddagger$  and  $\Delta G_d^\ddagger$ , respectivamente. Además, la evaluación de las constantes cinéticas en función de la temperatura permite ajustar los datos usando la forma lineal de la ecuación de Eyring:

$$\ln \left( \frac{k}{T} \right) = \ln \left( \frac{k_B}{h} \right) + \frac{\Delta S^\ddagger}{R} + \frac{\Delta H^\ddagger}{R} \cdot \frac{1}{T} \quad (9)$$

Siendo  $k$  la constante de velocidad ( $k_a$  o  $k_d$ ),  $k_B$ , la constante de Boltzman ( $1,38 \times 10^{-23} \text{ J} \cdot \text{K}^{-1}$ ),  $h$ , es la constante de Planck ( $6,626 \times 10^{-34} \text{ J} \cdot \text{s}$ ) y  $\Delta S^\ddagger$  y  $\Delta H^\ddagger$  corresponden a la entropía y entalpía de activación, respectivamente.

## 2.6. Desplegamiento al equilibrio seguido por dicroísmo circular.

Para los experimentos de despliegamiento al equilibrio las proteínas fueron preparadas en solución CD (20 mM  $\text{Na}_2\text{HPO}_4$ , pH 7 y 50 mM  $\text{Na}_2\text{SO}_4$ ) a una concentración de  $\sim 20 \mu\text{M}$  e incubadas con diferentes concentraciones de cloruro de guanidinio (GndCl; grado biología molecular, Thermo Fischer) entre 0 y 6 M. Todas las muestras fueron incubadas a  $37^\circ\text{C}$  por 2 h hasta que la reacción alcanzó el equilibrio. Las muestras fueron medidas en un espectropolarímetro Jasco J-1500 (Jasco) colectando el espectro entre 200 y 260 nm y

registrando la señal a 223 nm de cada muestra por duplicado. Las señales registradas fueron analizadas como se detalla en la siguiente sección.

## 2.7. Análisis del desplegamiento al equilibrio de las proteínas FoxP.

La concentración de proteína usada en los experimentos de desplegamiento al equilibrio en presencia de GndCl fue de  $\sim 20 \mu\text{M}$ . Los modelos que mejor explicaron el desplegamiento de las proteínas fue uno de dos estados ( $\text{N} \rightleftharpoons \text{2U}$ ), para FoxP2 y FoxP4, y uno de tres estados ( $\text{N} \rightleftharpoons \text{2I} \rightleftharpoons \text{2U}$ ) para FoxP3 y FoxP1. Donde N corresponde al estado nativo, I al intermediario, y U al estado desplegado. A diferencia del modelo de dos estados, donde el cambio de energía libre considera solo el desplegamiento del monómero, en el modelo de tres estados, el cambio en la energía libre debe considerar un evento de disociación ( $\Delta G_1$ ) y formación de un intermediario monomérico previo al desplegamiento ( $\Delta G_u = \Delta G_1 + 2 \Delta G_2$ ). Los cambios en la energía libre de ambos procesos se definen en las siguientes ecuaciones:

$$\Delta G_1 = RT \ln (K_1) \quad (10)$$

$$\Delta G_2 = RT \ln (K_2) \quad (11)$$

En donde  $K_1$  y  $K_2$  representan las constantes de equilibrio de las etapas de disociación y desplegamiento, respectivamente. Para estudiar la dependencia de los cambios de energía libre de Gibbs con la concentración de GndCl se utilizaron las siguientes ecuaciones:

$$\Delta G_1 = \Delta G_1^0 + m_1 [\text{GndCl}] \quad (12)$$

$$\Delta G_2 = \Delta G_2^0 + m_2 [\text{GndCl}] \quad (13)$$

En este caso,  $\Delta G_1^0$  y  $\Delta G_2^0$  representan los valores del cambio en energía libre Gibbs a una concentración de 0 M del agente desnaturante. Las pendientes de las respectivas rectas representan el cambio en la energía libre de disociación ( $m_1$ ) y desplegamiento del intermediario monomérico ( $m_2$ ).

La señal de los espectros de dicroísmo circular fue seguida a 223 nm ( $y_{\text{obs}}$ ) y esta fue asumida como la suma de las señales de la disociación y el desplegamiento del intermediario monomérico normalizada por las fracciones de cada una de las especies ( $y_{\text{obs}} = y_{\text{U}}f_{\text{U}} + y_{\text{N}}f_{\text{N}} +$

$y_I$   $f_i$ ). Las señales correspondientes a cada una de las especies varían de la misma forma que lo observado en la ecuación 12 y 13, lo que las define como sigue:

$$y_N = y_{N0} + m_n [\text{GndCl}] \quad (14)$$

$$y_I = y_{I0} + m_I [\text{GndCl}] \quad (15)$$

$$y_U = y_{U0} + m_U [\text{GndCl}] \quad (16)$$

De manera análoga a lo observado en las ecuaciones mencionadas anteriormente, las señales  $Y_{N0}$ ,  $Y_{I0}$  e  $Y_{U0}$ , representan el valor de las respectivas señales a una concentración de 0 M GndCl y las pendientes de esas rectas ( $m_n$ ,  $m_I$ ,  $m_U$ ) corresponden a la variación de las señales según sus respectivos estados (N, I, U).

## 2.8. Cristalización y difracción de cristales de FoxP4

Los experimentos de cristalización para FoxP4 fueron realizados en el equipo automatizado Gryphon (ART Robbins Instruments), utilizando una placa de 96 pocillos con condiciones de cristalización diferentes. FoxP4 fue preparada a una concentración de 7 mg/mL en un tampón HEPES pH 7,8, NaCl 150 mM,  $\beta$ -mercaptoetanol 2 mM. El método utilizado fue de difusión de vapor por gota sentada, con un volumen de 70  $\mu$ L de solución madre (solución precipitante) en el reservorio y una mezcla de 0,8  $\mu$ L de proteína y 0,8  $\mu$ L de solución madre. Se ensayó condiciones de cristalización con 5 *kits* y los cristales fueron crecidos a una temperatura de 17 °C. Los cristales fueron colocados en *cryo-loops* y preservados en frío en nitrógeno líquido.

Los experimentos de difracción de rayos-X fueron realizados en la línea MX2, del Laboratorio Nacional de la Luz Síncrotron (LNLS) del Centro de investigación en Energía y Materiales, Campinas, Brasil. Se utilizó un detector PILATUS 2M (Dectris) y una longitud de onda de 1,459 Å. Los datos colectados fueron indexados, integrados y escalados usando XDS (Kabsch et al., 2010) y mezcladas usando Aimless del paquete CCP4 (Evans & Murshudov, 2013). Las fases fueron resueltas usando el método de reemplazo molecular en Molrep (Vagin & Teplyakov, 2010) de CCP4-i2 (Potterton et al., 2018) utilizando como templado el monómero de FoxP2 (PDBid 2A07) (Stroud et al., 2006). El refinamiento

manual fue hecho en Coot (Emsley et al., 2010) y el automatizado en Phenix (Adams et al., 2010). El modelo y los datos fueron depositados en el Protein Data bank (PDB) con el código 6XAT y se encuentran retenidos hasta que se envíe una publicación derivada de esta tesis.

## **2.9. Análisis de identificación de residuos PCAIN.**

La identificación de residuos conservados en el núcleo hidrofóbico de proteínas fue realizada utilizando el lenguaje de programación TCL para la ejecución en la consola TK del programa VMD. Se utilizaron las proteínas monoméricas FoxO3 (PDBid: 2UZK(Tsai et al., 2007)) y FoxA2 (PDBid: 5X07 (Li et al., 2017) y las estructuras disponibles para FoxP2 (PDBid: 2AS5 (Wu et al., 2006), 2A07 (Stroud et al., 2006)) separando monómeros y dímeros con intercambio de segmentos. El *script* de análisis identifica residuos con baja accesibilidad al solvente (superficie accesible al solvente  $\leq 20 \text{ \AA}^2$ ) y luego a ese grupo de residuos seleccionados se le hace un análisis de contactos. Un contacto entre dos residuos se define cuando el centro de masa de las dos cadenas laterales tiene una distancia  $\leq 5 \text{ \AA}$ . El programa marca la columna beta de cada PDB y a partir de eso se analizan las estructuras y las posiciones específicas de cada proteína.

## **2.10. Simulaciones moleculares para evaluar el rol de la protonación de H59 de FoxP1**

Como se detalla en la publicación derivada de este trabajo de tesis (Medina et al., 2019), se realizaron simulaciones moleculares utilizando NAMD 2.9 (Phillips et al., 2005) con el campo de fuerza CHARMM36m (Huang et al., 2016). Se utilizó el monómero mutante de FoxP1 A39P/C61Y (PDBid: 2KIU (Chu et al., 2011)) el que fue mutado en VMD con un *script* en TCL/TK para restituir la secuencia de la proteína silvestre. Además, H59 fue preparada usando la topología de histidina protonada (HSP) y otra desprotonada (HSD). Cada proteína fue solvatada en una caja con moléculas de agua TIP3P y su carga neutralizada con iones Cl<sup>-</sup>. Los sistemas fueron minimizados usando el algoritmo de gradiente conjugado por 4000 pasos. Los sistemas fueron calentados a 300 K en pasos discretos de aumento de temperatura manteniendo la cadena principal restringida y todo el sistema fue equilibrado en un ensamble NPT por 20 ns. Se realizaron tres réplicas independientes por sistema (HSP y HSD) alcanzado 40 ns de fase de producción usando un paso de tiempo de 2 fs. La estabilidad conformacional de los sistemas fue evaluada calculando el RMSD.

La evaluación de la estabilidad de las trayectorias, visualización de trayectorias y análisis de puente de hidrógeno fue realizada en VMD (Humphrey et al., 1996). Para el análisis de puentes de hidrógeno entre H59 y residuos de la hélice H3 se utilizó la herramienta *Hydrogen Bonds* de VMD usando un criterio de distancia entre el dador y aceptor de hidrógeno de 3,0 Å y un ángulo de 20°.

### **2.11. Simulaciones moleculares de los dímeros y monómeros FoxP.**

Las simulaciones de monómeros (FoxP1, FoxP2, FoxP4) y dímeros FoxP (FoxP2, FoxP3) fueron realizados en AMBER18 (Case et al., 2005) usando el campo de fuerza ff14SB (Maier et al., 2015). La estructura de FoxP4 (PDBid: 6XAT) fue superpuesta en VMD/Multiseq (Phillips et al., 2005) con el monómero de FoxP2 (PDBid: 2A07) (Stroud et al., 2006) para añadir los residuos no resueltos en la hélice H5 de FoxP4. Los sistemas fueron preparados en *AmberTools*, neutralizando la carga con iones Cl<sup>-</sup> y se generó una caja de agua de 14 Å alrededor de la superficie de la proteína. Inicialmente, para acomodar el solvente y los átomos de hidrógeno agregados a la proteína, esta última fue restringida en fases iniciales del protocolo de minimización y equilibrado con una restricción harmónica (500 kcal/mol Å<sup>2</sup>). Las cadenas laterales de los residuos mutados o las regiones adicionales a la estructura de las proteínas como la hélice H5 en FoxP4 fue dejada sin restricciones para acomodarse al resto de la estructura durante las fases iniciales del protocolo de minimización y equilibrado. Los sistemas fueron minimizados con 1000 pasos de Descenso empinado y luego 1.000 pasos de gradiente conjugado. Los sistemas fueron calentados a 300 K y la fase final del equilibrado fue realizada en un ensamble NPT por 20 ns. Tres réplicas independientes de 180 ns fueron realizadas durante la fase de producción.

Las trayectorias fueron analizadas en CPPTraj de *AmberTools* para el cálculo de RMSD y radio de giro. El análisis para agrupar los clústeres más representativos fue realizado usando el algoritmo de *k-means* que fue configurado para obtener 10 clústeres y sus respectivas estructuras representativas (centroides). La matriz de correlación dinámica fue calculada usando el comando Matrix de CPPTraj.

## 2.12. Simulaciones de plegamiento de FoxP1

Las simulaciones moleculares que permitieron observar transiciones plegamiento-desplegamiento fueron realizadas con una aproximación simplificada de los átomos (*All-atom*) que simula todas las interacciones enlazantes con potenciales armónicos (ángulos, enlaces, diedros). Los contactos no enlazantes entre dos átomos se definen si estos están a una distancia menor a  $6,0 \text{ \AA}$  y no se encuentran apantallados por otro átomo entre ellos, considerando una distancia en la secuencia de al menos tres residuos (Noel et al., 2012)

Como estructura inicial fue usado un modelo de FoxP1 generado en Modeller v9.18 (Webb & Sali, 2014) la cual fue cargada al servidor SMOG para generar la estructura necesaria para la simulación y el respectivo campo de fuerza (Noel, et al., 2010). Las interacciones definidas por el servidor fueron simetrizadas en la subunidad B de modo de los contactos intra e inter-cadena en ambas subunidades del dímero fueran los mismos. Eso resultó en 433 contactos por monómero y 584 del dímero. Las simulaciones moleculares fueron realizadas en GROMACS 4.5.4 (Pronk et al., 2013) usando el integrador *Leap frog* para dinámica estocástica con un paso de tiempo de  $0,0002 \tau$  (en unidades reducidas) por  $1,6E8$  pasos a temperaturas entre 1,05 y 1,12 unidades reducidas de temperatura, próximas a la temperatura de plegamiento ( $T_F$ ) de FoxP1. Las estructuras de cada cuadro de la trayectoria y las energías fueron guardadas cada 2.000 pasos. Los paisajes energéticos fueron calculados usando el método de análisis de histogramas (WHAM) (Kumar et al., 1992a). Los contactos intra ( $Q_M$ ) e inter-cadena ( $Q_D$ ) fueron usados como coordenadas de reacción. Para la identificación de estos contactos en las trayectorias de FoxP1 se definió como el contacto entre dos residuos si es que estaban a 1,2 veces de su distancia nativa.  $Q_M$  y  $Q_D$  varían desde el estado desplegado (0) al plegado (1) y son obtenidos dividiendo el número de contactos observados para cada cuadro de la trayectoria por el número de contactos nativo.

### 2.13. Filogenia y reconstrucción de secuencias ancestrales.

Se descargaron 40 secuencias del dominio de unión a ADN de humanos de la familia Fox desde la base de datos UNIPROT (Bateman, 2019) y luego fueron alineadas usando el programa PRANK (Löytynoja, 2014). El modelo de sustitución de aminoácidos LG (matriz de reemplazo de aminoácidos General) (Le & Gascuel, 2008) fue determinado por Prottest (Darriba et al., 2011) como el mejor modelo que representa la evolución de este grupo de secuencias. La inferencia filogenética fue realizada la aproximación de máxima verosimilitud implementada en PhyML (Guindon et al., 2010), utilizado el alineamiento generando en PRANK y el modelo de sustitución LG. El soporte estadístico para cada rama del árbol fue calculado por PhyML usando el método de prueba de razón de verosimilitud aproximada (aLRT) (Anisimova & Gascuel, 2006).

La reconstrucción de las secuencias ancestrales más probables fue calculada en PAML 4.4 (Yang, 2007) incorporado en Lazarus (Hanson-smith et al., 2010). La reconstrucción ancestral estima qué aminoácido es más probable para cada posición en el alineamiento; como el alineamiento de la familia contiene *gaps*, esas posiciones también son reconstruidas y se estima los aminoácidos más probables para esa posición, resultando en una reconstrucción del largo total del alineaamiento. Para determinar qué posiciones corresponden a un *gap* o residuos en las secuencias ancestrales (nodos del árbol) calculadas por Lazarus/PAML, se representó el alineamiento de la familia como un modelo de rasgo evolutivo binario. Para esa labor se representa el alineamiento con solo dos caracteres: proteína (P) en los lugares donde se observa un aminoácido y *gap* (G) donde se observan espacios. Con ese alineamiento y el árbol filogenético se usa un modelo de rasgo evolutivo binario en el programa Mesquite (Mesquite Project Team, 2014) para estimar en qué posiciones de los nodos del árbol (ancestros) es más probable observar un aminoácido o un *gap*. La secuencia inicialmente estimada por Lazarus/PAML es editada según la información obtenida en Mesquite y con eso se obtuvo la secuencia más probable para cada ancestro. La secuencia calculada de AncP fue optimizada para uso de codones en *E. Coli* y sintetizada en GENSCRIPT (Piscataway), la sobreexpresión y purificación fue realizada según los protocolos descritos previamente.

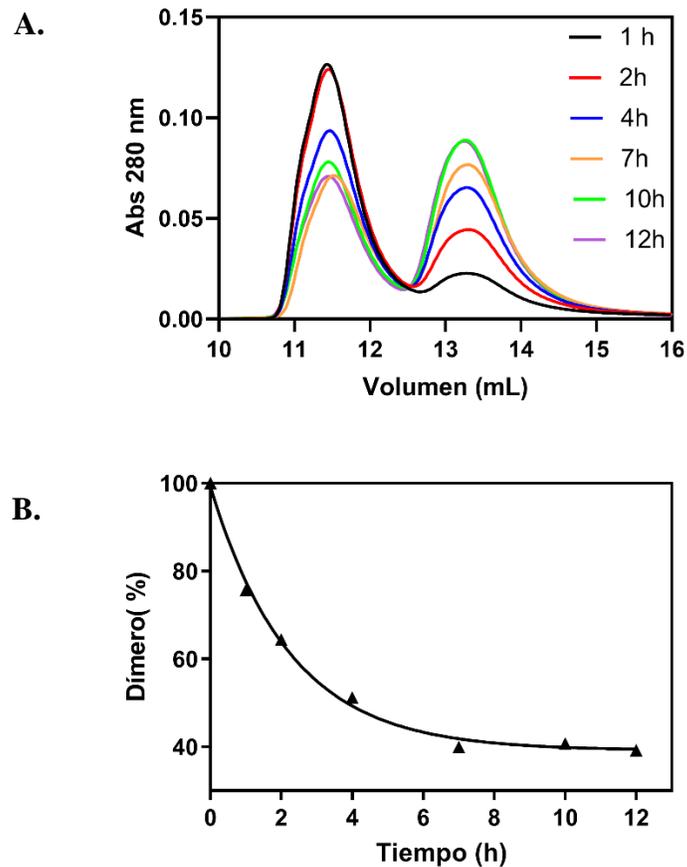
### 3. RESULTADOS.

#### 3.1. Cinéticas de disociación de FoxP4 y FoxP1.

Las pruebas experimentales estructurales y en solución de las proteínas de la subfamilia FoxP sugiere que la oligomerización por intercambio de segmentos es un mecanismo conservado en los miembros de esta familia (Bandukwala et al., 2011; Chu et al., 2011; Stroud et al., 2006). FoxP1 ha sido una de las proteínas más estudiada en esta subfamilia y de la cual se ha reportado que tiene una constante de disociación ( $K_D$ ) de orden  $\mu\text{M}$  bajo (Medina et al., 2016) (Chu et al., 2011)(Medina et al., 2019). Por el contrario, FoxP2 es la proteína con mayor tendencia a disociarse ( $K_D$  de orden mM) (Perumal et al., 2015) y que FoxP3 es un dímero obligado ( $K_D$  de orden nM) (Bandukwala et al., 2011), por lo cual la tendencia a la disociación de las proteínas FoxP varían significativamente con pocos cambios en la secuencia (identidad de secuencia promedio de 84%).

Sin embargo, aún se carece de estudios para FoxP4, el último miembro del que se desconocen sus propiedades de dimerización en la subfamilia. Para completar la caracterización biofísica de la subfamilia, se realizaron con esta proteína experimentos de cinética de disociación del dímero y se compararon con FoxP1, el miembro más caracterizado de la subfamilia FoxP.

Para caracterizar las propiedades de dimerización de FoxP4, el dímero fue aislado mediante cromatografía de exclusión molecular, y luego se siguió la cinética de disociación a partir de una concentración inicial de  $39 \mu\text{M}$  utilizando también cromatografía de exclusión molecular (Materiales y métodos), hasta que la reacción de disociación alcanzó el equilibrio (Figura 3 A). En el caso de FoxP1, la concentración de dímero usada en el ensayo fue de  $3 \mu\text{M}$ . Los datos obtenidos fueron ajustados a una exponencial de decaimiento simple, y se observó que el equilibrio de la reacción de disociación, para FoxP4, se alcanzó aproximadamente a las 10 h a  $25 \text{ }^\circ\text{C}$  (Figura 3 B). Como se mencionó anteriormente, FoxP1 se caracteriza por tener una baja  $K_D$ , lo que impide realizar este experimento a temperaturas bajo  $33 \text{ }^\circ\text{C}$  con las concentraciones de dímero requeridas para monitorear el proceso mediante absorbancia UV. En ese contexto, es que se realizaron los ensayos entre  $20\text{-}33 \text{ }^\circ\text{C}$  para FoxP4, mientras que para FoxP1 se realizaron entre  $33$  y  $44 \text{ }^\circ\text{C}$ .



**Figura 3. Cinética de disociación del dímero de FoxP4 a 25 °C.** **A.** Experimentos de cromatografía de exclusión molecular de FoxP4 realizados a diferentes temperaturas a una concentración de proteína de 39  $\mu$ M. La columna fue calibrada con un estándar de proteína Bio-Rad y se determinó que los picos observados a 11,5 y 13,5 mL corresponden al dímero y monómero de FoxP4, respectivamente. **B.** Cinética de disociación de FoxP4. Los picos del dímero en la figura A. fueron cuantificados y graficados en función del tiempo de incubación a 25 °C, luego esos datos fueron ajustados a una función exponencial de decaimiento simple como se detalla en Materiales y métodos.

Las constantes calculadas a partir de esos datos (Materiales y métodos) se muestran en la Tabla 1. A 33°C, la  $K_D$  de FoxP4 es de  $135 \pm 8.0 \mu\text{M}$ , mientras que FoxP1 posee una constante de  $1,0 \pm 0,20 \mu\text{M}$ , lo que en términos de energía libre indica que la disociación está favorecida en FoxP4 en  $\sim 3,0 \text{ kcal}\cdot\text{mol}^{-1}$  respecto a FoxP1. Considerando las constantes de disociación en la subfamilia FoxP, la  $K_D$  de FoxP4 se posiciona en un valor intermedio, entre las constantes calculadas para FoxP1 y FoxP2.

Las diferencias observadas en el equilibrio dímero-monómero de FoxP4 y FoxP1 pueden ser complementadas mediante el análisis de las constantes cinéticas de asociación ( $k_a$ ) o disociación ( $k_d$ ), lo que permite determinar la barrera energética que separa al monómero del dímero respecto al estado de transición. Con el ajuste a una función de decaimiento exponencial simple de las cinéticas de disociación del dímero se obtuvo el valor de la constante de cinética  $k_d$ , lo que permite, a partir de su relación con la  $K_D$ , obtener la constante cinética de asociación  $k_a$ , según la ecuación 6 (materiales y métodos). A 33 °C, la comparación de los valores cinéticos para FoxP1 y FoxP4 arrojó diferencias principalmente en la reacción de dimerización ( $k_a$ ), evidenciando que la asociación está desfavorecida para FoxP4 en  $\sim 25$  veces respecto a FoxP1, mientras que la disociación está favorecida solo  $\sim 6$  veces en FoxP4 respecto a FoxP1. Utilizando la aproximación de Eyring se determinó que la asociación está desfavorecida en  $\sim 1,8 \text{ kcal}\cdot\text{mol}^{-1}$  en FoxP4, respecto a FoxP1. Estos antecedentes sugieren que las diferencias observadas en las propiedades de dimerización de ambas proteínas están mayormente influenciadas por la asociación de los monómeros.

Para calcular la contribución entálpica al equilibrio de ambas proteínas, se hizo un análisis de van't Hoff observando la dependencia lineal de la  $K_D$  con la temperatura (Figura 4 A). Los valores de la diferencia de entalpía de disociación en condiciones de equilibrio ( $\Delta H_D$ ) estimados para FoxP1 y FoxP4 son  $17.2 \pm 2.7$  y  $10.5 \pm 1.5 \text{ kcal}\cdot\text{mol}^{-1}$ , respectivamente, indicando que la asociación de ambas proteínas está guiada por un componente entálpico y sugieren que este proceso involucra una ganancia de un número mayor de contactos entre los residuos de FoxP1 que estabilizan el dímero respecto a lo que ocurre con FoxP4. En adición al cálculo de la entalpía de disociación al equilibrio, se calcularon las barreras entálpicas de las cinéticas de disociación ( $\Delta H_d^\ddagger$ ) y asociación ( $\Delta H_a^\ddagger$ ). Para ello se observó la dependencia con la temperatura de las constantes cinéticas de disociación ( $k_d$ ) y asociación ( $k_a$ ) usando

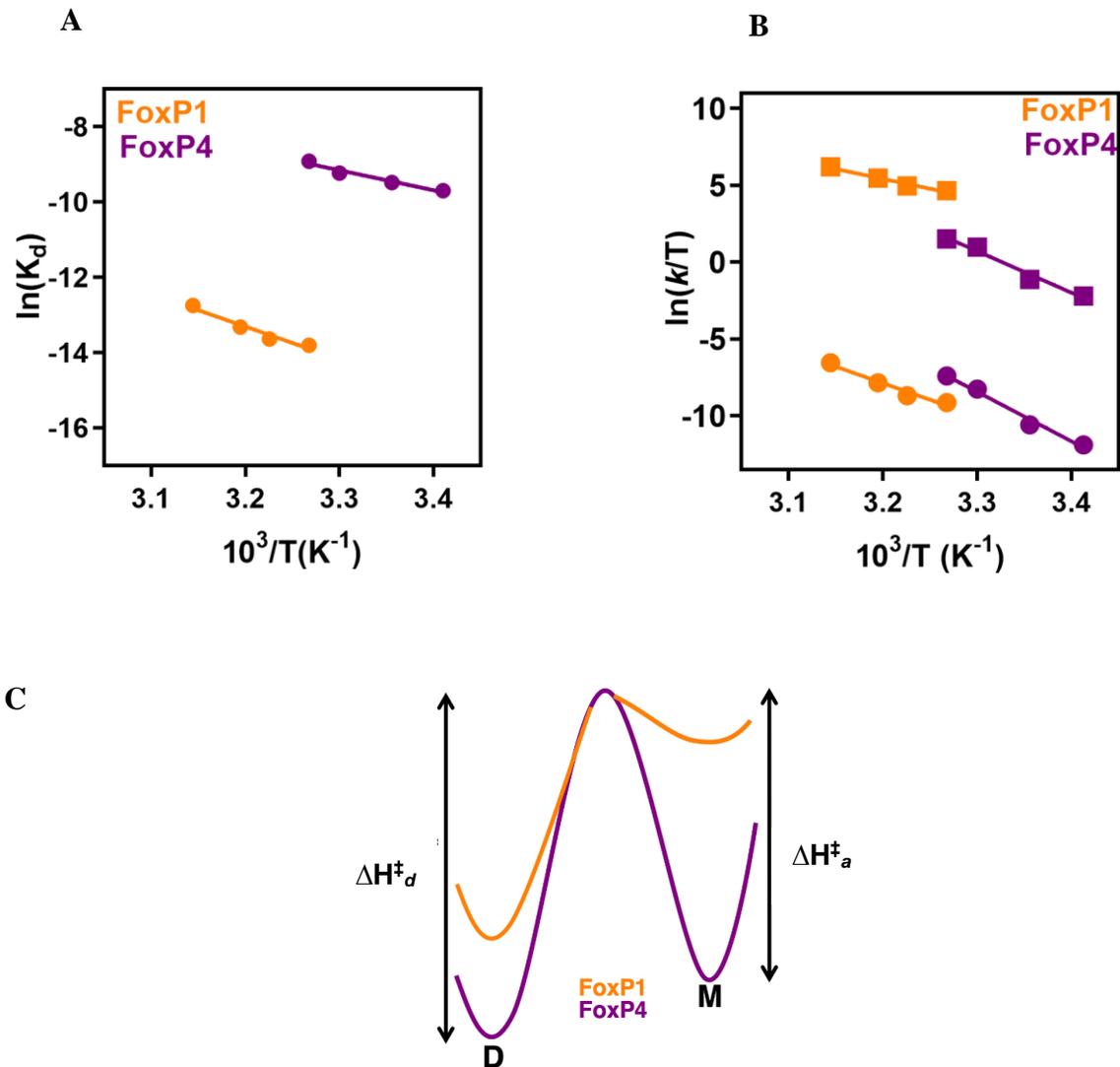
una ecuación de Eyring. Las barreras entálpicas calculadas para FoxP1 ( $\Delta H_a^\ddagger$ : 25,4±2,9 y  $\Delta H_a^\ddagger$ : 42,7±5,5 kcal·mol<sup>-1</sup>), en comparación a las calculadas para FoxP4 ( $\Delta H_a^\ddagger$ : 63.8±5.7 y  $\Delta H_a^\ddagger$ : 53.6±5.7 kcal·mol<sup>-1</sup>) indican que las diferencias en las propiedades de dimerización de ambas proteínas están dadas mayormente por su entalpía de asociación ( $\Delta H_a^\ddagger$ ) y, probablemente, sea aplicable para todos los miembros de la subfamilia debido a su alta similitud estructural y de secuencias.

**Tabla 1. Parámetros cinéticos y termodinámicos de las cinéticas de disociación del dímero de FoxP1 y FoxP4.**

T	FoxP1				FoxP4			
	$K_D$	$\Delta G_D^\S$	$k_d \times 10^3$	$k_a \times 10^{-3}$	$K_D$	$\Delta G_D^\S$	$k_d \times 10^3$	$k_a \times 10^{-3}$
(°C)	( $\mu M$ )		( $\text{min}^{-1}$ )	( $M^{-1} \cdot \text{min}^{-1}$ )	( $\mu M$ )		( $\text{min}^{-1}$ )	( $M^{-1} \cdot \text{min}^{-1}$ )
20	–	–	–	–	59,0±3,0	5,7±0,03	2,10±0,2	0,03±0,004
25	–	–	–	–	76,0±6,0	5,6±0,05	7,40±0,6	0,1±0,010
30	–	–	–	–	97,0±13	5,5±0,08	79,0±9,0	0,8±0,020
33	1,0±0,20	8,4±0,10	32,0±7,0	32,0±8,0	135±8,0	5,4±0,04	185,0±17	1,3±0,100
37	1,2±0,10	8,4±0,05	52,0±6,5	45,0±6,0	–	–	–	–
40	1,6±0,10	8,3±0,03	122,0±14,1	74,0±9,0	–	–	–	–
44	2,9±0,30	8,0±0,10	454,0±56,0	155,0±25,0	–	–	–	–

$\S$  kcal·mol<sup>-1</sup>

\*Las constantes y los respectivos errores provienen del ajuste de los datos de las cinéticas de disociación a una función de decaimiento exponencial, como se detalla en materiales y métodos.



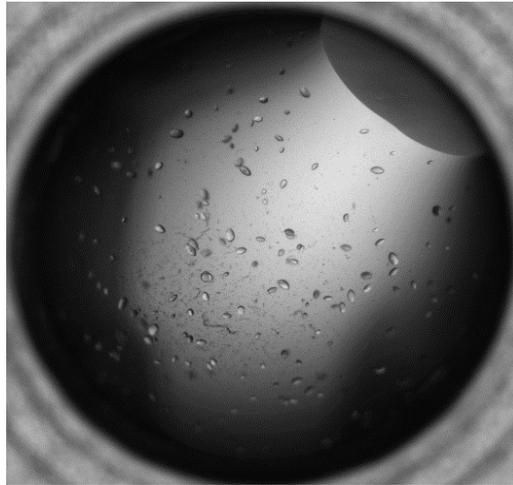
**Figura 4. Barrera entálpica de la transición dímoro monómero en FoxP1 y FoxP4.** La constante de disociación ( $K_D$ ) y las constantes cinéticas descritas en la Tabla 1 fueron usadas para observar su dependencia con la temperatura. **A.** Análisis de van't Hoff de dependencia de la  $K_D$  con la temperatura; a partir de la pendiente se obtuvo la contribución entálpica al equilibrio ( $\Delta H_D$ ) **B.** Dependencia de las constantes cinéticas de asociación (cuadrados) y disociación (círculos) con la temperatura. En naranja y morado se representan los datos de FoxP1 y FoxP4, respectivamente. Los datos fueron ajustados a la ecuación de Eyring y a partir de su pendiente se obtuvo la entalpía que separa al monómero ( $\Delta H_a^\ddagger$ ) y dímoro ( $\Delta H_d^\ddagger$ ) con el estado de transición. **C.** Se esquematizaron las diferencias en las barreras entálpicas derivadas de la dependencia de las constantes cinéticas con la temperatura.

### 3.2. Determinación de la estructura de FoxP4.

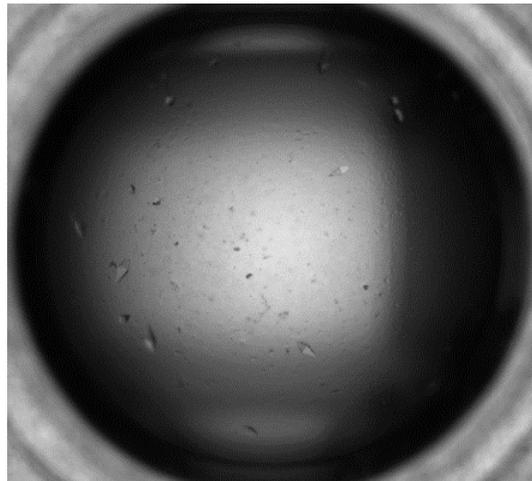
La caracterización cinética de la disociación de los dímeros de FoxP4 y FoxP1 sugiere que, respecto a FoxP1, la formación del dímero en FoxP4 está desfavorecida producto de la significativa mayor barrera energética de transición que lo separa del monómero. Respecto a la información estructural disponible para estas proteínas esta indica que las características de la subfamilia FoxP es su alta similitud estructural en el dominio de unión a ADN, tanto con el resto de la familia Fox como entre sus miembros. Lo anterior es sustentado por la comparación de las estructuras resueltas para FoxP3 (Bandukwala et al., 2011)(Chen et al., 2015), FoxP2 (Stroud et al., 2006) (Wu et al., 2006) y un monómero de la mutante A39P/Y61C de FoxP1 (Chu et al., 2011). El último miembro sin estructura conocida es FoxP4, por lo que se realizaron ensayos de cristalización para intentar obtener su estructura en ausencia y presencia de ADN.

Para los ensayos de cristalización de FoxP4 se utilizó una concentración de proteína de 7 mg/ml en tampón HEPES 20 mM a pH 7,8, siendo la máxima concentración que pudo ser alcanzada sin observar agregación de proteína, lo cual es necesario para la obtención de cristales. Los experimentos permitieron encontrar dos condiciones en donde se observaron cristales: en ausencia y presencia de ADN. Mientras que para la condición de cristalización en ausencia de ADN involucra 4 M de formato de sodio, la condición en presencia de ADN corresponde a una solución de citrato trisódico dihidratado 0,1 M pH 5,0 y jeffamida-2001 pH 7,0. Los cristales de proteínas para ambas condiciones crecieron en aproximadamente dos semanas a 25 °C y su tamaño se encontró en el rango de los 20-50  $\mu\text{m}$ , lo que los califica como cristales de tamaño pequeño (Figura 5). Los cristales fueron sometidos a difracción de rayos-X, obteniéndose patrones de difracción para cuatro cristales en ausencia de ADN y solo en uno en presencia de este. Luego del procesamiento de los patrones de difracción se determinó que dos de los cristales difractados poseían una resolución máxima de 2,2 Å y 3,5 Å, correspondiendo a la condición en ausencia y con ADN, respectivamente. Sin embargo, el proceso de refinamiento no pudo ser llevado a cabo para la proteína en presencia de ADN ya que la baja resolución y ausencia de densidad electrónica clara no permitió resolver esta estructura. Por otra parte, la estructura en ausencia de ADN fue resuelta utilizando como modelo para el reemplazo molecular el monómero de FoxP2 (cadena K, PDBid:2A07) (Figura 6). La información de la colección de datos y el proceso de refinamiento se encuentra

**A**



**B**



**Figura 5. Cristales de proteína de FoxP4 en ausencia y presencia de ADN.** Luego de la purificación de FoxP4, la proteína fue concentrada a una concentración final de 7 mg/mL en tampón Hepes pH 7,8, con lo que fue sometida a ensayos de cristalización en un robot de cristalización ARI Gryphon, utilizando la aproximación de difusión de vapor por gota sentada. **A.** Cristales de FoxP4 en ausencia de ADN y crecidos en presencia de formato de sodio 4M **B.** Cristales de FoxP4 crecidos en presencia de un análogo de ADN y en solución de tri-sodio citrato dihidratado 0,1M pH 5,0 y jeffamida-2001 pH 7,0.

resumida en la Tabla 2. Uno de los parámetros usados para evaluar la calidad de un modelo es el factor R, que es un parámetro que muestra qué tan bien se ajusta el modelo los datos experimentales. Un valor de 0 indica una perfecta representación de los datos experimentales en el modelo y un valor cercano a 0,60 indica una distribución de los átomos al azar en el modelo. En este caso, y a la resolución observada de 2,2 Å, el factor R alcanzó un valor de 0,20, lo que está dentro de lo observado para estructuras en ese rango de resolución (Kleywegt & Brünger, 1996). Sin embargo, el factor R puede ser disminuido incrementando el número de parámetros para ajustar el modelo durante el proceso de refinamiento. Por lo mismo, otro valor que se reporta junto al factor R es el factor R libre, que indica qué tan bien se ajusta el modelo a un conjunto de datos reservados y que no son usados en el proceso de refinamiento, por lo que si el modelo es correcto es esperable tener valores siempre superiores al factor R pero cercanos a éste (Kleywegt & Brünger, 1996). En efecto, el valor obtenido de R libre para este modelo fue de 0,22, sugiriendo un buen ajuste del modelo a esos datos. La estructura fue validada y depositada en el banco de datos de proteínas (PDB) con el código 6XAT y se encuentra retenida para su liberación luego de la publicación de un artículo producto de esta tesis.

El análisis de la estructura resuelta en ausencia de ADN muestra un monómero para FoxP4 (Figura 6 A) con un átomo de sodio coordinado por las cadenas laterales de S56, H59, L55, F62 entre la hélice H3 y la hebra S2. El monómero conserva las características típicas del plegamiento en la subfamilia FoxP, tal como se observa en la estructura cristalográfica para FoxP2 (Stroud et al., 2006) y una mutante monomérica (A39P/C61Y) resuelta por RMN de FoxP1 (Chu et al., 2011). La única diferencia respecto a los otros monómeros resueltos en la subfamilia es que la estructura de FoxP4 no tiene densidad electrónica suficiente para resolver la estructura de la hélice H5 en el extremo carboxilo terminal, por lo que el modelo final no incluye este elemento de estructura secundaria. Mientras que la presencia de esta hélice en las estructuras de la subfamilia FoxP ha sido resuelta, en el resto de la familia Fox también hay casos donde esa estructura no es parte del modelo, como en FoxO1 (PDBid: 3CO7) (Brent et al., 2008) y FoxA2 (PDBid: 5X07)(Li et al., 2017). Además, en otros miembros como FoxM1 (PDBid: 3G73) (Littler et al., 2010) o FoxC2 (PDBid: 6AKP)(Chen et al., 2019), esta región o parte de ella aparece desestructurada, lo que explicaría la alta flexibilidad y por lo tanto falta de densidad electrónica para FoxP4.



**Figura 6. Estructura resuelta del monómero de FoxP4 y superposición estructural de los monómeros de la subfamilia FoxP.** A. Representación en cintas de la estructura cristalográfica del monómero de FoxP4. La cadena principal está representada en color morado y se destaca en amarillo un átomo de sodio. Los residuos que se encuentran a 5 Å del átomo de sodio se representan con varas. Superposición estructural hecha en VMD de las estructuras de FoxP4 (PDBid: 6XQT, morado), mutante monomérica de FoxP1 (A39P/C61Y, PDBid: 2KIU; naranja) y FoxP2 (PDBid: 2A07, verde). S56, H59, L55, F62 entre la hélice H3 y la hebra S2.

**Tabla 2. Parámetros de recolección de datos y refinamiento del monómero de FoxP4.**

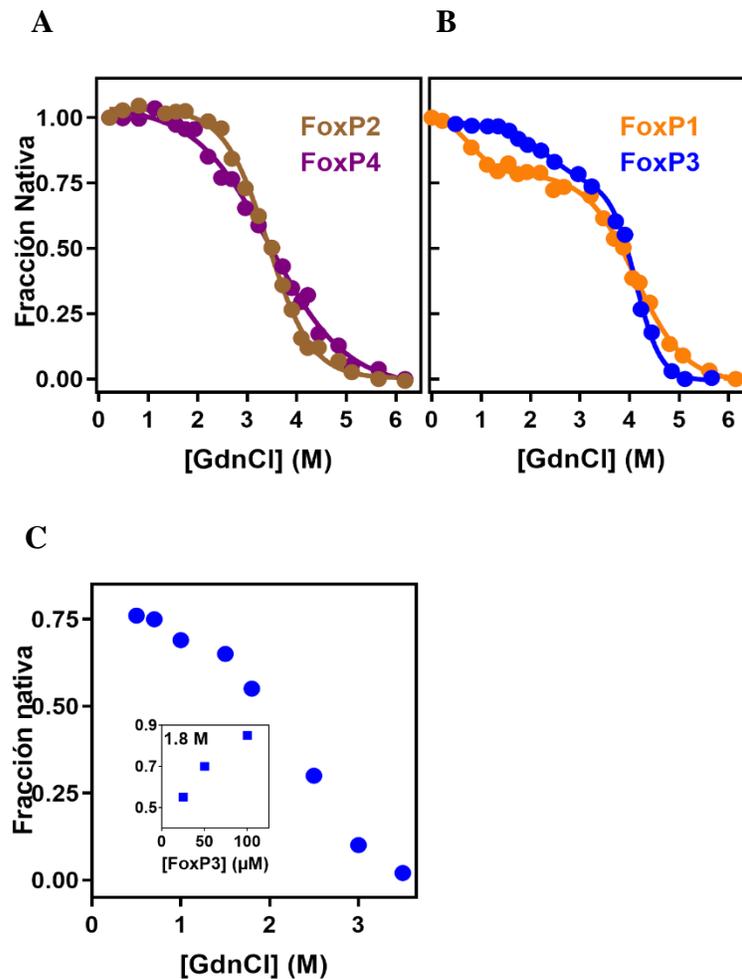
<b>FoxP4</b>	
<b>Recolección de Datos</b>	
Fuente de rayos-X	LNLS-MX2
Detector	Pillatus 2M
Longitud de onda(Å)	1,45869
Rango de resolución	35,1 - 2.2 (2,3 – 2,2)
Grupo Espacial	C 2 2 21
Celda Unitaria (a b c, $\alpha$ $\beta$ $\gamma$ )	40,5 70,7 63,5, 90 90 90
Reflexiones	60496 (6039)
Reflexiones únicas	4862 (488)
Multiplicidad	12,4 (12,4)
Complejidad (%)	99,90 (100,00)
I/sigma(I)	24,70 (2,53)
Wilson B-factor	42,17
R-merge	0,075 (0,977)
R-pim	0,022(0,286)
CC1/2	1 (0,81)
<b>Refinamiento</b>	
R /R libre	0,20/0,22
<b>Número de Átomos</b>	
Proteína	645
Agua	21
<b>Factores B</b>	
Proteína	46,04
Agua	49,29
<b>Desviaciones promedio</b>	
Enlaces	0,020
Ángulos	1,31
<b>Gráfico de Ramachandran</b>	
Favorecidos (%)	98,68
Permitidos (%)	1,32
No permitidos (%)	0

Las estructuras disponibles para los monómeros de la subfamilia FoxP provienen de condiciones diferentes: FoxP2 y FoxP4 son estructuras cristalográficas, donde la primera fue cristalizada en presencia de ADN, mientras que la segunda en ausencia del oligonucleótido. Por otra parte, la estructura resuelta para FoxP1 fue obtenida por resonancia magnética nuclear (RMN) y es una mutante (A39P/C61Y) monomérica (Chu et al., 2011), en cuyo caso ambas mutaciones se localizan separadas entre sí en la estructura: una en la región bisagra (A39P) y otra en la hebra S3 (C61Y). Sorprendentemente, a pesar de las diferentes condiciones en que cada estructura fue obtenida, la superposición estructural de la cadena principal de los tres monómeros FoxP demuestran que no hay grandes variaciones ni cambios conformacionales globales, incluso en una proteína mutante y monomérica como la reportada para FoxP1. Esto se ve reflejado en cómo la superposición estructural de FoxP4 con FoxP1 y FoxP2 resulta en un RMSD de 1,2 Å y 0,7 Å, respectivamente (Figura 6 B). Este antecedente sugiere que las diferencias en las propiedades de oligomerización de la subfamilia FoxP no involucran cambios conformacionales importantes a nivel de la cadena principal en el monómero en la subfamilia, por lo que es pertinente ahondar en otros aspectos más allá de la estructura que permitan comprender la diferencias en el mecanismo de intercambio de segmentos de segmentos

### 3.3. Estabilidad y mecanismo de plegamientos en la subfamilia FoxP.

Los antecedentes expuestos anteriormente muestran que los procesos de asociación y disociación en la subfamilia FoxP están fuertemente gobernados por los pocos cambios en la secuencia de cada proteína en la subfamilia sin tener un impacto en la estructura nativa de los monómeros pese a las importantes diferencias en las barreras cinéticas de asociación observadas en FoxP1 y FoxP4, como referencia. Sin embargo, se ha propuesto como mecanismo general de intercambio de segmentos que el desplegamiento total de los monómeros y el replegamiento a concentraciones elevadas de proteína es necesario para alcanzar el dímero entrelazado (Rousseau et al., 2001)(Liu et al., 2012). Por ello, este proceso ha sido descrito como un mecanismo de dos estados ( $N_2 \rightleftharpoons 2U$ ) sin la presencia de intermediarios observables con las técnicas utilizadas. Contrario a lo establecido como mecanismo general, la evidencia con FoxP1 indica que el mecanismo de dimerización y plegamiento no involucra el tránsito por el estado desplegado para alcanzar el dímero entrelazado (Medina et al., 2016), sino que más bien por un estado intermediario (I) parcialmente desplegado del monómero, lo que ha sido descrito como un mecanismo de tres estados ( $N_2 \rightleftharpoons 2I \rightleftharpoons 2U$ ). Pese a la alta similitud de secuencia y estructural de las proteínas FoxP respecto a otras proteínas de la familia Fox, no existen antecedentes que indiquen si este mecanismo es conservado en sus miembros. Considerando el contraste entre la conservación de secuencia y estructural de las proteínas FoxP y sus diferentes propiedades de dimerización, es atractivo estudiar las posibles vías de plegamiento que puedan explicar la formación del dímero con intercambio de segmentos.

Para determinar si la estabilidad termodinámica de las proteínas FoxP ha modelado a lo largo de la evolución la oligomerización por intercambio de segmentos en la subfamilia FoxP, se realizaron experimentos de desplegamiento al equilibrio de sus miembros, siguiendo la reacción por medio de la cuantificación del contenido de estructura secundaria de las proteínas a una temperatura de 37 °C. Para fines comparativos se utilizó una concentración de 20  $\mu$ M por lo que, dadas las constantes de disociación reportadas en este estudio, el dímero está altamente favorecido en FoxP1 y FoxP3, mientras que el monómero es el estado favorecido a esta concentración de proteína para FoxP2 y FoxP4. Los resultados se muestran en la Figura 7 y de acuerdo con lo anterior, se apreciaron dos patrones claros de desplegamiento: una única transición a (2,5-4,5 M GndCl) y dos transiciones (0,5-1M y 2,5-



**Figura 7. Despliegamiento al equilibrio de las proteínas FoxP.** Las proteínas, a una concentración de  $\sim 20 \mu\text{M}$ , se incubaron a diferentes concentraciones de agente desnaturalante (GdnCl), siguiendo los cambios en la elipticidad a 223 nm en experimentos de dicroísmo circular. **A.** Los datos obtenidos para FoxP2 y FoxP4 fueron ajustados a un modelo de despliegamiento de dos estados **B.** Los datos obtenidos para FoxP3 y FoxP1 fueron ajustados a un modelo de tres estados, el que conecta el estado nativo con el desplegado por medio de un intermediario monomérico de plegamiento. **C.** Se incubaron muestras del dímero de FoxP3 con diferentes concentraciones de desnaturalante y la fracción del dímero fue cuantificada calculando el área bajo la curva en experimentos de cromatografía de exclusión molecular. El gráfico inserto muestra la cuantificación de la población dimérica a una misma concentración de agente desnaturalante (1,8 M) y diferentes concentraciones de proteína.

4,5 M de GndCl), correspondiendo el primer patrón a FoxP2 y FoxP4 y el segundo a FoxP1 y FoxP3.

De acuerdo con lo que se ha publicado (Medina et al., 2016) y por el estado de agregación esperado según las concentraciones utilizadas en estos experimentos, el modelo que explica el desplegamiento de FoxP3 y el de FoxP1 corresponde a un mecanismo de tres estados, con la acumulación de un intermediario monomérico ( $N_2 \rightleftharpoons 2I \rightleftharpoons 2U$ ) que retiene cerca de un 80% del contenido de estructura secundaria del estado nativo y que conecta al estado Nativo (N) y desplegado (U). La presencia de este intermediario monomérico ha sido descrita anteriormente en FoxP1 (Medina et al., 2016), pero no en FoxP3, a pesar de que es altamente probable que exista en esta última, dada las similitudes en las propiedades de dimerización y la identidad de secuencia que ellas comparten. Por lo anterior, se hizo un ensayo de exclusión molecular donde se observó tanto la variación de la población dimerica de FoxP3 a diferentes concentraciones de agente desnaturante, como el efecto de la concentración de proteína en la primera transición (0,5-1M de GndCl). Los resultados mostraron, por un lado, que la población de dímero decrece fuertemente con la concentración de agente desnaturante entre 0 y 3 M de GndCl (Figura 7 C), rango en el cual se observa solo un 20% de cambio en la señal dicróica, y por otro, que este proceso es dependiente de la concentración de proteína (inserto en Figura 7 C). Esto último sugiere que, al igual que lo propuesto para FoxP1, el desplegamiento de FoxP3 involucra la disociación de las cadenas de ambas subunidades estableciendo un intermediario monomérico, y que puede ser descrito por un mecanismo de tres estados como el propuesto para FoxP1. Finalmente, y de manera contraria a lo expuesto anteriormente, el desplegamiento de FoxP2 y FoxP4 es descrito por un modelo de dos estados ( $N \rightleftharpoons U$ ), lo que es consistente con la concentración de proteína usada, ya que se encuentra dos y cuatro órdenes de magnitud debajo la  $K_D$ , respectivamente. Además, el mismo mecanismo fue utilizado para describir el desplegamiento de la mutante monomérica de FoxP1 (A39P) (Medina et al., 2016), por lo que estos antecedentes sugieren que cuando el monómero está favorecido, las proteínas FoxP siguen un mecanismo de desplegamiento de dos estados, el cual probablemente es un mecanismo conservado a lo largo de la toda la familia Fox.

Los parámetros energéticos obtenidos a partir de los ajustes de los desplegamientos de las proteínas FoxP se muestran en la Tabla 3. Interesantemente, el cambio en energía libre de desplegamiento ( $\Delta G_u$ ) es notoriamente diferente entre las cuatro proteínas, lo que sugiere que poseen estabilidades diferentes en las condiciones ensayadas: por una parte, FoxP1 y FoxP3, que en la concentración de proteína utilizada para estos experimentos se encuentran como dímero en solución, tienen un  $\Delta G_u$  de  $19,5 \pm 0,41$  y  $30,1 \pm 1,5$  kcal/mol, respectivamente. En contraste, FoxP2 y FoxP4, que en las condiciones ensayadas se encuentran como monómeros en solución, tienen un  $\Delta G_u$  de  $5,0 \pm 0,28$  y  $2,7 \pm 0,19$  kcal/mol, respectivamente. Si se considera únicamente a las especies monoméricas (monómero nativo de FoxP2-4 y el estado monomérico intermediario de FoxP1 y FoxP3) para la comparación de las diferencias en  $\Delta G_u$  entre las distintas proteínas, se observa que FoxP3 el intermediario es  $\sim 1,8$  veces más estable que FoxP1,  $\sim 2$  veces más estable que FoxP2 y  $\sim 3,6$  veces más estable que FoxP4.

Teniendo en cuenta que FoxP3 ha sido caracterizada como un dímero obligado y que FoxP1 tiene una constante de disociación del orden micromolar, estos resultados sugieren que la presencia de ese intermediario monomérico altamente estructurado está correlacionada con la capacidad de dimerizar de estas proteínas. Así, las diferencias en el mecanismo de plegamiento de los miembros de la subfamilia FoxP muestran que el desplegamiento parcial de estas proteínas está estrechamente relacionado con sus propiedades de dimerización. En efecto, las notables diferencias entre la estabilidad de FoxP1 y FoxP3, las dos proteínas con mayor tendencia a dimerizar de la subfamilia, pueden atribuirse, en base a lo expuesto aquí, a la capacidad de asociación de los monómeros producto de la presencia de un intermediario monomérico altamente estructurado.

**Tabla 3. Parámetros termodinámicos de los experimentos de desplegamiento al equilibrio de las proteínas FoxP.**

	FoxP1			FoxP2	FoxP3			FoxP4
	N <sub>2</sub> ⇌ 2I	I ⇌ U	N <sub>2</sub> ⇌ 2U	N ⇌ U	N <sub>2</sub> ⇌ 2I	I ⇌ U	N <sub>2</sub> ⇌ 2U	N ⇌ U
$\Delta G_u^{\S}$	8,5±0,50	5,5±0,40	19,5±0,8	5.0±0,3	11,0±1,6	9,8±1,1	31,0±2.2	2,7±0,2
$m$ - value <sup>¶</sup>	2,8±0,04	1,3±0,10	5,5±0,2	1.4±0,1	2,1±0,7	2,4±0,3	6,9±0.8	0,8±0,05

<sup>§</sup> kcal·mol<sup>-1</sup>

<sup>¶</sup> kcal·mol<sup>-1</sup>·M<sup>-1</sup>

\*Las constantes y sus respectivos errores fueron obtenidos a partir de los ajustes de dos y tres estados descritos en materiales y métodos.

### **3.4. Análisis de cambios en las redes de interacción en el núcleo hidrofóbico de las proteínas FoxP en el estado monomérico y dimérico.**

La comparación estructural de los monómeros FoxP sugiere que no hay cambios significativos a nivel de cadena principal que permita entender las diferencias en oligomerización entre los distintos miembros de la subfamilia FoxP. En ese sentido, se hace pertinente investigar si es que los contactos entre los residuos que estabilizan a las estructuras de estas proteínas en el estado monomérico y dimérico puede dar indicios a nivel de la estructura nativa de cuáles pueden ser las determinantes estructurales de estas diferencias.

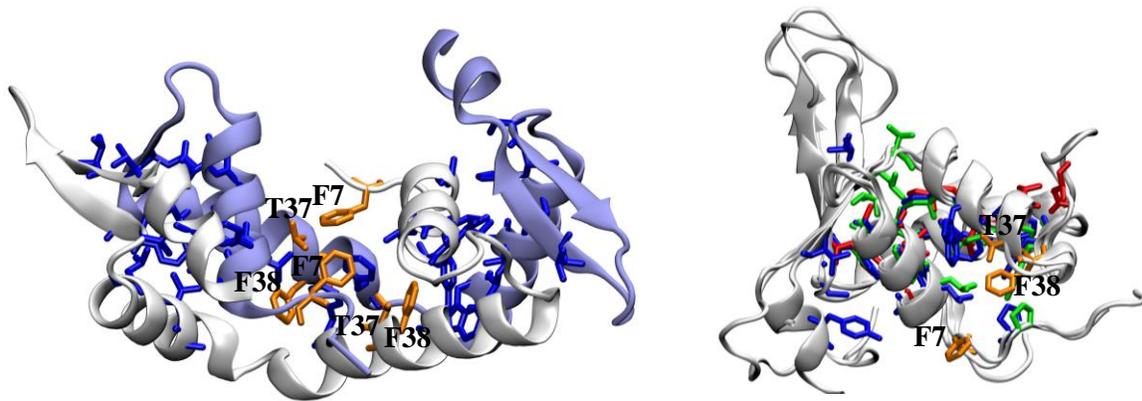
Por definición, el intercambio de segmentos es un mecanismo de oligomerización donde los contactos estabilizan al monómero (contactos intra-subunidad) deben romperse para generar una superficie de interacción que permita formar los mismos contactos, pero con la otra subunidad (contactos inter-subunidad), estabilizando el dímero con intercambio de segmentos (Yanshun & Eisenberg, 2002). En consecuencia, cuando los contactos de las estructuras de monómeros y dímeros se comparan, y apegándose a la definición de intercambio de segmentos, es esperable tener las mismas interacciones entre residuos tanto en el estado monomérico como dimérico. Sin embargo, las pruebas presentadas hasta aquí muestran que algunas proteínas FoxP tienen una probabilidad mayor de encontrarse como monómeros y otras como dímeros a concentraciones de orden micromolar. Por ello, es esperable encontrar dos tipos de interacciones al comparar la estructura de un monómero con un dímero formado por este mecanismo: aquellas que se mantienen entre ambos estados de oligomerización (responsables del intercambio de segmentos) y otras que están solo en uno de los dos estados de agregación, favoreciendo el equilibrio de dimerización hacia uno u otro estado.

Para analizar los cambios conformacionales en las cadenas laterales de las proteínas FoxP se hizo un análisis con un *script* implementado en TLC/TK (VMD) (Humphrey et al., 1996) que permite identificar residuos en las redes de interacciones atómicas en el núcleo de proteínas (PCAIN: *Protein Core Atomic Interaction Networks*) (Soundararajan et al., 2010). El objeto de identificar redes de interacción en el núcleo hidrofóbico de las proteínas es determinar aquellos residuos que son importantes para mantener la estructura nativa de la proteína analizada. Para eso, se buscan en las estructuras aquellos residuos de baja exposición al solvente y que tengan al menos un contacto con otro aminoácido. El análisis se ejecutó

usando como representante de la subfamilia FoxP a FoxP2, porque es la única proteína que tiene estructura conocida para el monómero y el dímero con intercambio de segmentos en la subfamilia FoxP, permitiendo comparar ambos estados de agregación y determinando de esta forma qué residuos son importantes para uno u otro estado.

Para ello se utilizó la información estructural disponible para FoxP2 PDBid: 2A07 (Stroud et al., 2006), considerando para el análisis las cadena de un dímero y un monómero de esa estructura. Además, para encontrar un patrón estructural que diferencie a las proteínas FoxP con el resto de la familia Fox monomérica, en la búsqueda de residuos responsables del origen del intercambio de segmentos en la familia, se utilizó como representantes las estructuras de FoxO3 (PDBid: 2UZK(Tsai et al., 2007)) y FoxA2 (PDBid: 5X07 (Li et al., 2017) )

Los resultados del análisis identificaron residuos PCAIN en todo el núcleo hidrofóbico de FoxO3 y FoxA2 (Figura 8). Sin embargo, su comparación con el monómero de FoxP2 no sugiere un patrón de conservación que permita proponer residuos que diferencien a las proteínas monoméricas de la familia Fox con aquellas que pueden oligomerizar por intercambio de segmentos como FoxP2 (Figura 8). Por otra parte, como era de esperar, y ya que por definición los contactos entre el dímero y monómero se deben conservar en proteínas que oligomerizan por intercambio de segmentos, muchos residuos identificados como PCAIN en el dímero también lo son en el monómero (cadena lateral azul). Sin embargo, la comparación de los residuos PCAIN entre las cadenas polipeptídicas de los monómeros y dímeros mostró que solo cuatro residuos fueron clasificados como PCAIN exclusivamente en el dímero (Figura 8, cadena lateral coloreada en naranja), localizándose mayormente en la hélice H1 y parte de la hélice H2 (región bisagra). Esto quiere decir que esos residuos sufren cambios conformacionales que permiten su participación en redes de interacción con otros residuos en el dímero, mientras que en el monómero no. Entre los residuos identificados como exclusivos del estado dimérico se encuentran T37 y S21 de la hélice H2, cuyas cadenas polares pueden establecer interacciones vía puentes de hidrógeno con otros residuos a nivel intra y/o inter-subunidad.



**Figura 8. Redes de interacción inter-residuo en el núcleo hidrofóbico de las proteínas FoxP.** Se identificaron los residuos con baja exposición al solvente y aquellos que participan en contactos con otros a una distancia menor a 5 Å. **A.** Dímero de FoxP2 (PDBid: 2A07) los residuos identificados como PCAIN están coloreados en azul y naranja. En naranja se muestran aquellos residuos que fueron identificados como PCAIN exclusivamente en las cadenas polipeptídicas del dímero de FoxP2. Las cadenas polipeptídicas de FoxP2 fueron coloreadas en azul y blanco. **B.** Superposición estructural de los monómeros de FoxP2 (PDBid:2A07), FoxO y FoxA. Los residuos PCAIN identificados en el monómero de FoxP2 están coloreados en azul, y de acuerdo con la definición de intercambio de segmentos, se encuentran conservados en dímeros y monómeros. Solo tres residuos fueron identificados como PCAIN en el dímero y no en el monómero, cuyas cadenas laterales fueron coloreadas en naranja.

Sin embargo, los cambios más notables se encuentran en los residuos F7 (hélice H1) y F38 (Hélice H2, región bisagra), que en el monómero de FoxP2 orientan sus cadenas laterales hacia el solvente, mientras que en el dímero aparecen formando una red de interacciones que estabilizan la interfaz entre ambas hélices. Estas interacciones en el dímero se suman a otras compuesta por varios residuos, principalmente aromáticos, provenientes de las hélices H1, H2 y H3, y que fueron identificados también como PCAIN tanto para el estado monomérico como dimérico. Esta compleja red de interacciones sugiere que, para que el dímero logre formarse durante la asociación, no solo deben reestablecerse los contactos observados en el monómero, sino que, también, deben formarse otros contactos entre ambas cadenas polipeptídicas, los que han permitido estabilizar un dímero con intercambio de segmentos durante la evolución de la subfamilia FoxP.

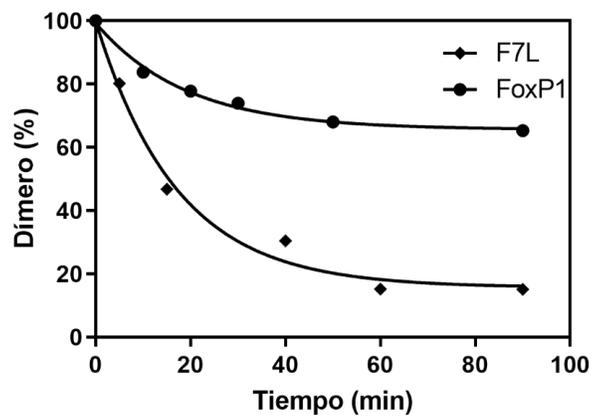
La red de interacciones que se observa en el dímero de FoxP2, establecida principalmente por residuos aromáticos entre las hélices H3, H2 y H1, también se observa en el dímero de FoxP3. Por lo anterior, y considerando que ambas proteínas tienen propiedades de oligomerización muy diferentes, es esperable que varios de los contactos que facilitan la asociación de los monómeros en la familia FoxP sean conservados en torno a esta interfaz en toda la subfamilia FoxP. Como parte de esa red, se puede destacar un residuo de tirosina o leucina en la posición 7 en el resto de la familia Fox (proteínas monoméricas), que es reemplazada por un residuo de fenilalanina propia de todos los miembros de la subfamilia FoxP. Basándose en ese aspecto, la presencia de fenilalanina y no tirosina o leucina pareciera ser una característica de aquellas proteínas que pueden oligomerizar en la familia Fox. En ese sentido, de acuerdo con los resultados de los análisis estructurales de residuos PCAIN, la presencia de una fenilalanina sugeriría que este residuo podría favorecer la desestabilización del monómero y ser una parte importante de la red de interacciones entre cadenas laterales aromáticas que se forma durante la asociación de los monómeros y que estabilizan al dímero con intercambio de segmentos.

### 3.5. Caracterización de una mutante PCAIN.

Para investigar el efecto de las redes establecidas por los residuos PCAIN, se diseñó y caracterizó la mutante F7L en FoxP1. F7 es un residuo conservado en todas las proteínas FoxP, mientras que en la misma posición en el resto de la familia Fox se observa leucina o tirosina. Se eligió leucina como reemplazo para fenilalanina, ya que la subfamilia de FoxO tiene leucina en esa posición. Con la introducción de un residuo de cadena lateral hidrofóbica pero voluminosa, como leucina, se esperaba desestabilizar la red de interacciones aromáticas que se observan en el dímero de FoxP2 y FoxP3, afectando la dimerización en FoxP1.

La proteína mutante fue caracterizada siguiendo la cinética de disociación de su dímero a 37 °C, de similar forma a lo realizado para las proteínas silvestre de FoxP1 y FoxP4 (Figura 9). Interesantemente, las diferencias entre FoxP1 silvestre y la mutante F7L fueron importantes, ya que la  $K_D$  de la mutante aumentó cerca de 24 veces respecto a FoxP1. Esas diferencias en la  $K_D$  se ven reflejadas en términos de la energía libre de la disociación al comparar el  $\Delta G_D$  de la proteína silvestre con las mutante, indicando que la disociación está favorecida en F7L en cerca de  $2,0 \text{ kcal}\cdot\text{mol}^{-1}$  respecto a FoxP1 (Tabla 4). El análisis de las barreras energéticas que separan al dímero del monómero en la mutante F7L muestran que esta mutante afecta la barrera energética de la asociación ( $\Delta G_a^\ddagger$ ) mayormente, aumentándola respecto a la proteína silvestre y de esta forma desfavoreciendo la formación del dímero con intercambio de segmentos.

Esto sugiere que el anillo aromático de F7 y los contactos que forma en la interfaz inter-cadena son importantes para la formación del dímero con intercambio de segmentos. Esto también es extrapolable al resto de las proteínas FoxP, debido a que esta fenilalanina está conservada en todos los miembros. Respecto a los otros residuos PCAIN identificados en la sección anterior, es muy probable que también tengan un impacto similar, ya que todos forman contactos inter-subunidad. Además, este resultado también da indicios de que el intercambio de segmentos puede ser modulado por residuos que se encuentran fuera de la región bisagra, la cual contiene la alanina que ha sido propuesta como la responsable del origen de intercambio de segmentos en esta subfamilia. En este escenario, estos cambios en oligomerización causados por la mutación de residuos PCAIN proveen una explicación a las propiedades diferentes de oligomerización de los miembros de esta subfamilia.



**Figura 9. Cinética de disociación del dímero de las mutantes PCAIN.** El dímero aislado de las proteínas fue aislado y luego se cuantificó su variación en el tiempo a 37 °C, siguiendo su absorbancia a 280 nm en experimentos de cromatografía de exclusión molecular. Los datos fueron ajustados a una función de decaimiento exponencial simple. Los círculos corresponden a la cinética de FoxP1 silvestre y los rombos a la mutante PCAIN F7L.

**Tabla 4. Parámetros cinéticos y termodinámicos de la disociación de la mutante PCAIN F7L.**

	<b>K<sub>D</sub></b> ( $\mu$ M)	<b><math>\Delta</math>G<sub>D</sub></b> ( $\mu$ M)	<b><math>\Delta</math>G<sup>‡</sup><sub>d</sub></b> kcal·mol <sup>-1</sup>	<b><math>\Delta</math>G<sup>‡</sup><sub>a</sub></b> kcal·mol <sup>-1</sup>
FoxP1	1,2±0,20	8,4±0,10	22,5±0,10	14,1±0,10
F7L	28,9±7,70	6,4±0,16	22,4±0,11	16,0±0,19

\*Las constantes y los respectivos errores provienen del ajuste de los datos de las cinéticas de disociación a una función de decaimiento exponencial, como se detalla en materiales y métodos.

### **3.6 Modulación del intercambio de segmentos por una histidina conservada en las familias FoxO, FoxM y FoxP.**

La hélice H3 es la responsable de la unión a ADN (Stroud et al., 2006)(Bandukwala et al., 2011)(Medina et al., 2016) en la familia Fox y, por lo tanto, estudiar la relación de ésta con el mecanismo de intercambio de segmentos es atractivo para relacionar la oligomerización con la función de estas proteínas. En efecto, se ha determinado que el pH afecta la unión a ADN en FoxP2, y ese efecto ha sido atribuido a la protonación de H54 (Blane & Fanucchi, 2015) de la hélice H3, siendo conocido el rol de este residuo que se inserta en la surco mayor del ADN (Stroud et al., 2006). Sin embargo, en la región del carboxilo terminal de esta hélice se localiza otra histidina (H59) (Figura 10 A), la que se encuentra conservada no solo en la subfamilia FoxP, sino que también en las subfamilias FoxO y FoxM. Por el contrario, en esa posición el resto de la familia tiene un residuo conservado sin carga como asparragina, sugiriendo que la presencia de esta carga es una novedad evolutiva en la familia Fox, por lo que puede ser una mutación que afecte la dimerización por intercambio de segmentos.

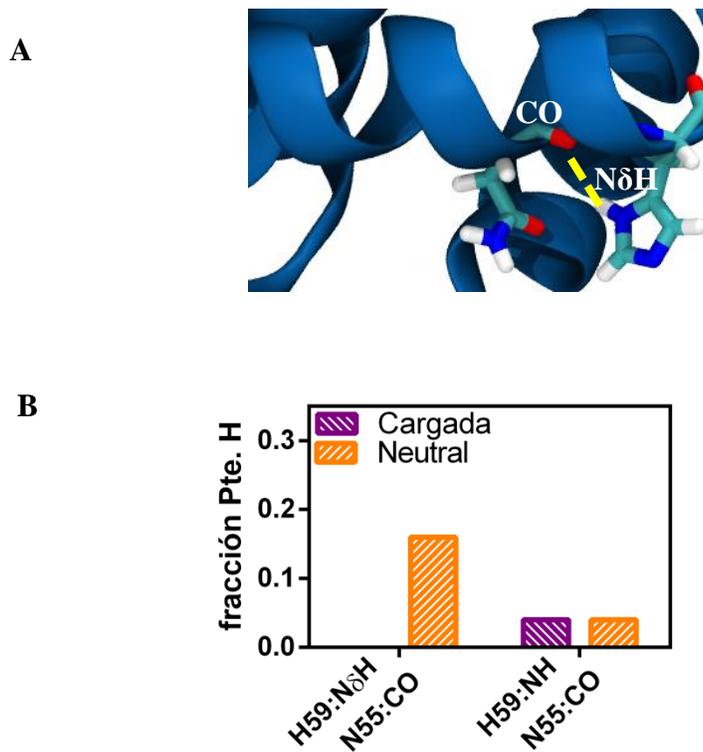
Recientemente nuestro grupo de investigación publicó un trabajo en donde se estudia el efecto del estado de protonación de la H59 en la estabilidad y en la dimerización de FoxP1 (Medina et al., 2019). En ese estudio se realizaron experimentos de cinética de disociación del dímero, demostrando que la protonación de este residuo incrementa la barrera energética entre el monómero y el estado de transición que lo separa del dímero. Interesantemente, el efecto fue causado mayormente en el proceso de asociación, lo que sugiere que la protonación de este residuo afecta mayormente al monómero y no al dímero. Adicionalmente, en experimentos de anisotropía en donde se marcaron la hélice H5 y H3 con fluoróforos para seguir el desplegamiento local de estos elementos en el contexto de la proteína completa, se demostró que el estado de protonación de H59 tiene un efecto específico en la estabilidad de la hélice H3.

Para evaluar el efecto molecular del estado de protonación de H59 en FoxP1, y como parte de los resultados de esta tesis, en esta publicación (Medina et al., 2019) se realizaron simulaciones moleculares cortas (40 ns) simulando el estado desprotonado (HSD) y protonado de la histidina (HSP) de FoxP1 monomérica. Con esto, se buscó observar el efecto del estado de protonación de la histidina sobre la estructura y dinámica conformacional del

monómero, ya que experimentalmente los efectos se observaron en el proceso de asociación. Las simulaciones se hicieron usando el programa NAMD (Phillips et al., 2005) con ambos sistemas embebidos en una caja de agua y realizando tres réplicas para cada uno.

En estas simulaciones moleculares no se observó un efecto importante a nivel de la estructura global del monómero de FoxP1 como consecuencia del estado de protonación de este residuo. Sin embargo, el análisis local de las interacciones alrededor de H59 sugiere que la protonación de la histidina desestabiliza la región del carboxilo terminal de la hélice H3. Este efecto se observa principalmente por la perturbación de la interacción de H59 con N55, siendo este último un residuo muy conservado en la mayoría de las Fox humanas monoméricas, lo que sugiere que esta interacción puede ser importante para la estabilidad del monómero. En ese sentido, y como las interacciones formadas por puente de hidrógeno son claves para la estabilización de los elementos de estructura secundaria en las proteínas (Bordo & Argos, 1994)(Hubbard & Haider, 2010), se analizó su formación en las simulaciones moleculares.

Para eso se utilizó la herramienta *Hydrogen bonds* de VMD, cuantificando los puentes de hidrógeno con un criterio de distancia ( $\leq 3.0 \text{ \AA}$ ) y de ángulo de formación ( $\leq 20^\circ$ ) entre el dador y aceptor de hidrógeno. El análisis mostró que el anillo imidazol de H54 desprotonada interactuó durante el 16% de la simulación con el grupo carbonilo del enlace peptídico del residuo N54, mientras que no se observó dicha interacción en la simulación con la histidina protonada (Figura 10 B). Respecto a las interacciones entre los grupos amino y carbonilo de la cadena principal, no se observaron diferencias dependientes del estado de protonación de H59, lo que indica que el efecto de la protonación de este residuo es directo entre su cadena lateral y la cadena principal de N55. Estos resultados fueron complementados con simulaciones moleculares de plegamiento del monómero de FoxP1, que mostraron que la desprotonación de H59 aumenta la probabilidad de contacto de la interacción H59-N55, así como también la de contactos entre otros residuos de la hélice H1 y hebra S2 (Medina et al., 2019).



**Figura 10. Formación de puentes de hidrógeno entre H59 y N55 como consecuencia de la desprotonación de la histidina.** **A.** Interacción H59-N54 en la hélice H3. Se destaca con líneas punteadas en amarillo el puente de hidrógeno entre la cadena lateral de la H59 neutral (HSD) y grupo carboxilo de N54. **B.** Cuantificación de los puentes de hidrógeno entre la cadena lateral de H59 (NδH) y la cadena principal de N55 (CO). Adicionalmente, se cuantificaron los puentes de hidrógeno establecidos entre sus respectivas cadenas principales (NH y CO), sin observar una diferencia importante dependiente de la protonación de la histidina.

Los resultados obtenidos en este trabajo de tesis y que forman parte de la publicación anteriormente mencionada, sugieren que el anillo imidazol de histidina, localizado en una región alejada de la bisagra, es capaz de modular el intercambio de segmentos en FoxP1. Este efecto probablemente es extrapolable al resto de la subfamilia FoxP, ya que la composición aminoacídica en la hélice H3 de sus miembros es muy similar y la histidina se encuentra conservada en todos.

### 3.7. Análisis de las mutantes derivadas de la comparación de secuencias FoxP.

La comparación de estructuras, secuencias y mecanismo de oligomerización de los miembros de la subfamilia FoxP muestran que, a lo largo de la evolución, han ocurrido pequeños cambios en sus secuencias que han derivado en propiedades de oligomerización diferentes entre ellas. Las comparaciones a nivel de estructura y secuencia permiten sugerir que existen cambios a estos dos niveles que pueden dar cuenta de las diferentes propiedades de oligomerización en la familia Fox. De acuerdo con lo anterior, se propusieron mutaciones en FoxP1 para favorecer la dimerización, usando como modelo para esas mutaciones los aminoácidos en posiciones equivalentes en la secuencia de FoxP3, que constituye un dímero obligado en esta subfamilia. Por otra parte, basándose en que FoxP2 y FoxP4 tienen una constante de disociación más alta que FoxP1, se usó las secuencias de ambas proteínas como modelos para diseñar mutantes que favorezcan la disociación en FoxP1.

Para favorecer la dimerización en FoxP1, se diseñaron tres mutantes en base a la presencia de residuos que se observan en la secuencia de FoxP3 y no en las posiciones equivalentes de la secuencia de los otros miembros de la subfamilia FoxP (Figura 2). Las mutantes diseñadas involucran distintas regiones de la proteína: Y41F en la región bisagra, Q15W localizada en la hélice H1 y V69E en la aleta (W1) (Figura 11 A). La localización de F41 en la región bisagra de FoxP3 es clave, ya que está ubicada a dos residuos de A39, la mutante responsable del origen del intercambio de segmentos en la familia (Stroud et al., 2006). Además en la posición equivalente de FoxP2 se realizó la mutación Y40F, resultando este cambio en una disminución de la  $K_D$  en cerca de 9 veces (Perumal et al., 2015). W15 ha sido reportado como un residuo único en FoxP3 dentro de la subfamilia, el que conforma en la estructura cristalográfica una interfaz secundaria de esta proteína y que ha sido propuesta como la responsable de conferir las propiedades de dímero obligado de FoxP3 (Bandukwala et al., 2011). La mutación S20A se localiza en una zona desestructurada entre H1 y S1, en la cual FoxP3 tiene una cadena hidrofóbica como alanina y el resto de la familia una cadena lateral polar sin carga como treonina o serina. Finalmente, E69 se localiza en la aleta W1 y en la estructura aparece expuesto al solvente, sin establecer contactos con otros aminoácidos en el resto de la estructura, siendo este uno de los pocos cambios únicos de FoxP3 respecto al resto de la subfamilia en el extremo carboxilo terminal de la proteína, ya que gran parte de

la variabilidad de secuencia en estas proteínas se da en la región del amino terminal (H1-S1-H2).

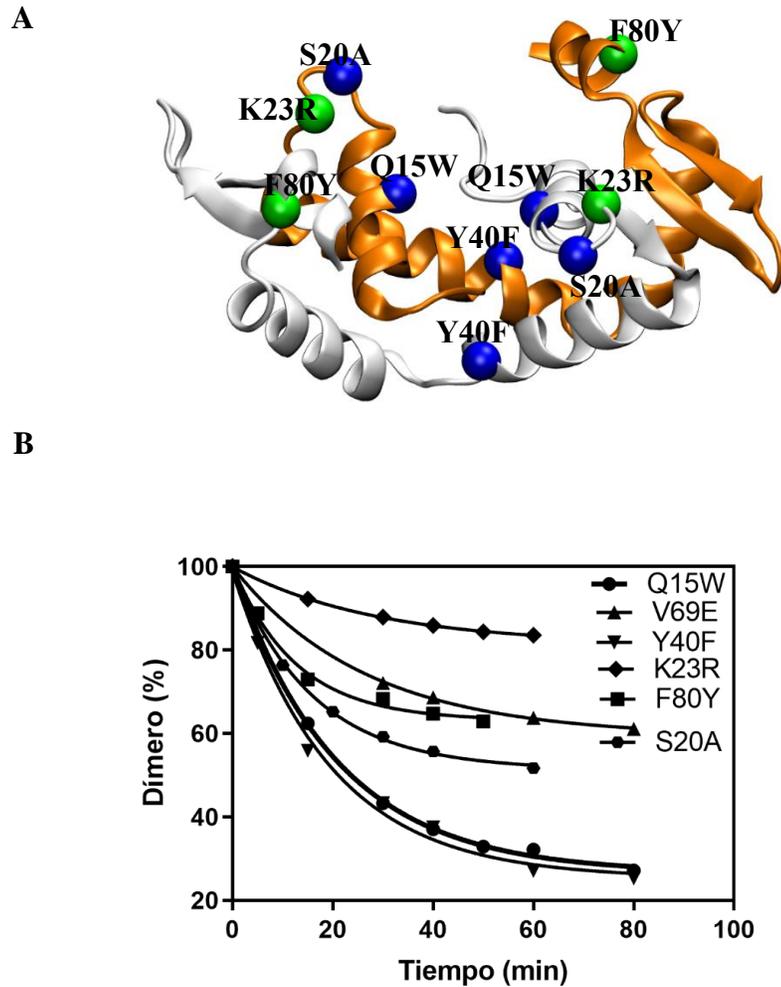
Para evaluar el efecto de esas mutantes, se realizaron experimentos de cinéticas de disociación del dímero, cuyos resultados se muestran en la Figura 11 A. El cálculo de la constante de la disociación permitió observar que, contrario a lo esperado, todas las mutantes aumentan la constante de disociación respecto a la enzima silvestre (Tabla 5). La mutación S20A aumentó dos veces la constante de disociación, V69E la aumentó aproximadamente cuatro veces, por lo que ambas mutaciones generan un efecto leve en la proteína. Mientras que Y40F y Q15W la aumentaron 27 y 26 veces, respectivamente. El efecto de las Y40F y Q15W se explica mayormente por un aumento en la barrera energética de asociación ( $\Delta G_a^\ddagger$ ), que separa al monómero con el estado de transición, aumentándola en ambas proteínas en  $\sim 2$  kcal·mol<sup>-1</sup>, mientras que la barrera energética de disociación ( $\Delta G_d^\ddagger$ ) se mantiene en todas las mutantes muy próxima a la de FoxP1 ( $22,5 \pm 0,10$  kcal·mol<sup>-1</sup>) (Tabla 5).

Para desfavorecer la dimerización en FoxP1, se diseñaron dos mutantes usando como referencias los aminoácidos equivalentes en FoxP2 y FoxP4, dos proteínas con una constante de disociación considerablemente mayor que la de FoxP1. Ambas mutaciones se localizan en diferentes regiones de la estructura: F80Y en la hélice H5 y K23R en una región desestructurada entre la hélice H1 y S1 (Figura 11 A). La hélice H5, donde se ubica F80, se encuentra formando contactos con la hélice H1, que estabilizan la interacción entre ambas hélices, por lo que es esperable que la incorporación de tirosina en esa posición, como se observa en FoxP4 y FoxP2, pueda favorecer la formación de puentes de hidrógeno a través del grupo hidroxilo de la cadena lateral con la misma hélice H1. Por otra parte, la perturbación de la región en que se localiza K23 puede interferir en los contactos que durante la asociación deben romperse para reestablecer los contactos inter-cadena en el dímero, esto porque se ubica en una región desestructura justo al extremo amino terminal de la hebra S1, que en el dímero se contacta con la hebra S2 de la otra subunidad (Figura 11).

Las mutantes mencionadas fueron caracterizadas siguiendo la cinética de disociación del dímero y siendo comparadas con FoxP1. Sorprendentemente, en este caso también las mutantes incrementaron las constantes de disociación, pero en mucha menor magnitud que las mutantes basadas en FoxP3. Los cambios en  $K_D$  fueron mínimos, lo que se evidencia en

un incremento de tres veces la  $K_D$  para K23R, mientras que F80Y disminuye la constante de disociación a la mitad de FoxP1 (Tabla 5). Esta pequeña modificación de la  $K_D$  también puede ser explicado por un leve incremento en la barrera cinética de asociación ( $\Delta G_a^\ddagger$ ), por lo que se puede concluir que ambas mutaciones no modifican mayormente las propiedades de dimerización de FoxP1

Los resultados expuestos hasta aquí sugieren que los cambios de secuencia que han permitido a las proteínas de la subfamilia FoxP diferenciarse una de la otra no pueden explicarse por cambios en residuos puntuales entre una secuencia y otra. Aparentemente, debido a que las mutaciones propuestas han involucrado diversas regiones de las proteínas, incluso alejadas de la zona de la bisagra, los cambios que le confieren la particularidad en las propiedades de oligomerización tremendamente diferentes se deben a efectos sinérgicos que involucran más de un residuo y que durante la evolución de estas proteínas probablemente ocurrieron de forma concertada.



**Figura 11. Cinética de disociación del dímero de las mutantes de FoxP1.** Estructura de FoxP1 que muestra la ubicación de las mutaciones propuestas para modificar el equilibrio dímero-monómero. **A.** Las mutantes diseñadas en base a la comparación de secuencias para desplazar el equilibrio dímero-monómero hacia el dímero tienen el  $\alpha$  coloreado en azul (basadas en la secuencia de FoxP3), mientras que las mutantes para desplazar el equilibrio hacia el monómero se muestran en verde (basadas en las secuencias de FoxP2 y FoxP4). **B.** Todas las mutantes fueron caracterizadas a 37 °C a una concentración cercana a 3  $\mu$ M. Las líneas continuas corresponden al ajuste de los datos a una ecuación de decaimiento exponencial simple. La constante cinética  $k_d$  y las constantes derivadas a partir de ella como se detalla en materiales y métodos se muestran en la Tabla 5.

**Tabla 5. Parámetros cinéticos y termodinámicos de disociación del dímero de las mutantes de FoxP1.**

	$K_D$	$\Delta G_D$	$\Delta G_d^\ddagger$	$\Delta G_a^\ddagger$
	( $\mu M$ )	$kcal \cdot mol^{-1}$	$kcal \cdot mol^{-1}$	$kcal \cdot mol^{-1}$
FoxP1	1,2 $\pm$ 0,20	8,4 $\pm$ 0,10	22,5 $\pm$ 0,10	14,1 $\pm$ 0,10
Q16W	31,2 $\pm$ 1,30	6,4 $\pm$ 0,26	22,5 $\pm$ 0,02	16,2 $\pm$ 0,03
Y40F	32,8 $\pm$ 3,90	6,3 $\pm$ 0,07	22,5 $\pm$ 0,01	16,2 $\pm$ 0,01
V69E	4,2 $\pm$ 0,11	7,6 $\pm$ 0,01	22,7 $\pm$ 0,02	15,0 $\pm$ 0,02
F80Y	3,0 $\pm$ 0,22	7,8 $\pm$ 0,04	22,2 $\pm$ 0,10	14,4 $\pm$ 0,10
K23R	0,5 $\pm$ 0,01	8,9 $\pm$ 0,01	22,7 $\pm$ 0,02	13,9 $\pm$ 0,01
S20A	2,4 $\pm$ 0,10	7,9 $\pm$ 0,02	22,4 $\pm$ 0,03	14,4 $\pm$ 0,03

\*Las  $K_D$  y los respectivos errores provienen del ajuste de los datos de las cinéticas de disociación a una función de decaimiento exponencial, como se detalla en materiales y métodos.

### 3.8. Dinámica conformacional de la familia FoxP.

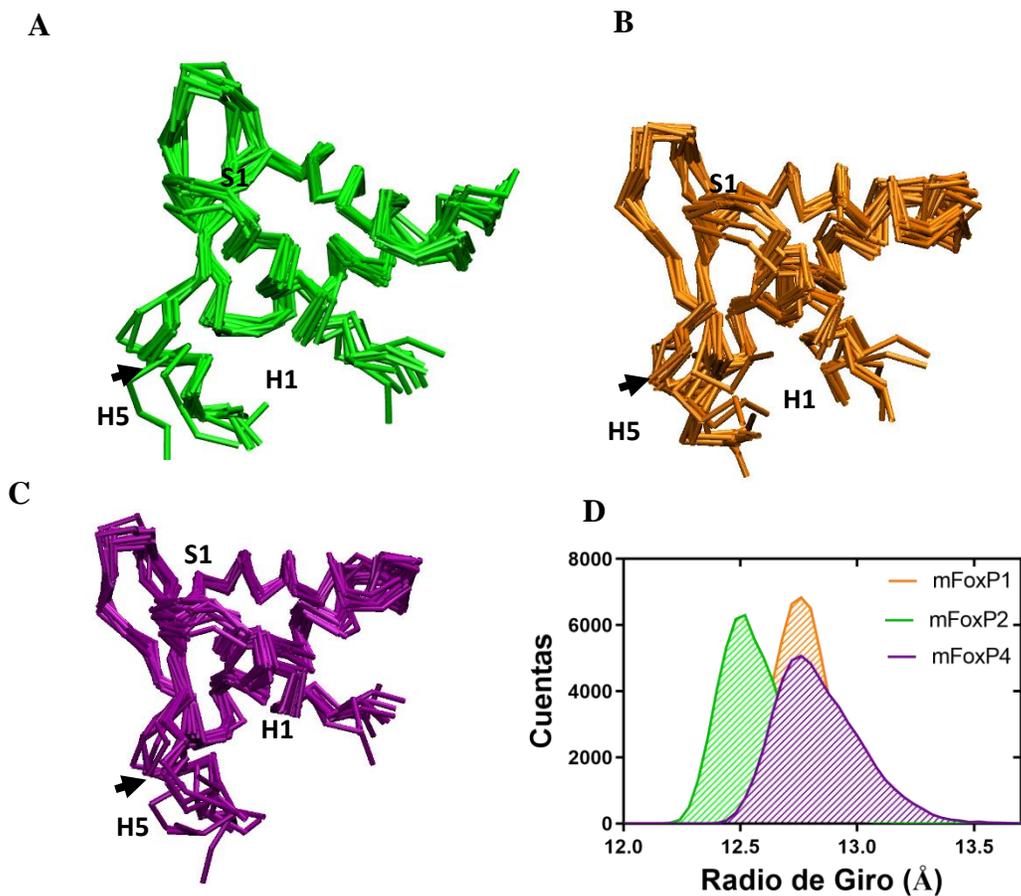
Los resultados mostrados hasta aquí sugieren que la oligomerización por intercambio de segmentos requiere que eventos de desplegamiento parcial en vez que un desplegamiento total de las proteínas para formar el dímero. Sorprendentemente, la comparación de la cadena principal de las estructuras de los monómeros no da cuenta, a nivel de estructura, de las diferencias observadas en las propiedades de oligomerización. Por esta razón, se estudió la dinámica conformacional de las proteínas FoxP, realizando simulaciones moleculares convencionales a partir de las estructuras disponibles en el banco de datos de proteínas (PDB). En particular se buscó determinar si los mínimos cambios en secuencia entre los miembros de la subfamilia FoxP dan cuenta de diferencias en la dinámica conformacional de estas proteínas, que actúen de manera sinérgica en modular la estabilidad del estado dimérico.

Todas las estructuras de los monómeros utilizados contemplaron 79 residuos por subunidad y en el caso de la región del carboxilo terminal de FoxP4, ésta fue modelada mediante superposición estructural con FoxP2. El único cambio de secuencia en ambas proteínas fue completado mediante la herramienta de mutagénesis de PyMol (DeLano, 2002) así como también para efectuar los cambios necesarios en la estructura mutante de FoxP1 (P39A y C61Y). Las proteínas fueron solvatadas en una caja de agua de 14 Å y sus cargas netas, las cuales fueron positivas para todos los sistemas de simulación, fueron neutralizadas con iones cloruro. Todos los sistemas fueron minimizados y equilibrados por 20 ns y su fase de producción fue llevada a cabo por 180 ns en triplicados.

Para analizar las trayectorias generadas se hizo un análisis de clústeres para las tres proteínas, de modo de obtener estructuras representativas para observar cambios conformacionales a lo largo de la simulación. Los *clusters* de FoxP2 la muestran manteniendo una conformación similar a la de la estructura cristalográfica, excepto dos, con una representatividad baja del 3% del total de la trayectoria (*clusters* 8 y 9, Tabla 6) en donde la hélice H5 se separa de H1. La región más variable conformacionalmente durante las simulaciones en FoxP4 y FoxP1 también es la hélice H5 (Figura 12 B y 12 C). Esas diferencias se aprecian en el análisis de la distribución del radio de giro, el que muestra que las proteínas tienen un grado de compactación diferente (Figura 12 D). FoxP2 mostró un máximo en su distribución de los radios de giro ( $R_g$ ) de 12,50 Å, mientras que FoxP4 y

FoxP1 fluctúan en una conformación muy similar con un valor máximo en su distribución de radio de giro ( $R_g$ ) de 12,75 Å (Figura 12 D). Además, FoxP4 tiene una conformación expandida, pero con una baja representación en las trayectorias (Tabla 6, *cluster 7*). Esto indica que, durante las simulaciones, el monómero de FoxP2 es más compacto que los monómeros de FoxP1 y FoxP4, lo que se debe a que mantiene la conformación cristalográfica (Figura 12 A). En el caso de FoxP1 y FoxP4, se observa una mayor variabilidad conformacional de la hélice H5 (Figura 12 B y 12 C), respectivamente, mientras que este mismo elemento de estructura secundaria permanece estable en FoxP2 dada la mantención de contactos con la hélice H1 presentes en la estructura cristalográfica a lo largo de la simulación (Figura 12 A). Consistentemente, la inestabilidad conformacional de esa hélice en FoxP1 y FoxP4 se debe a la pérdida de interacciones que estabilizan la conformación observada en las estructuras de los monómeros. Esta interfaz formada entre las hélices H1 y H5 está estabilizada por un puente salino entre R14 (hélice H1) y E77 (hélice H5) que, además, está estabilizado por una interacción de  $\pi$ -stacking entre F/Y80 y R14.

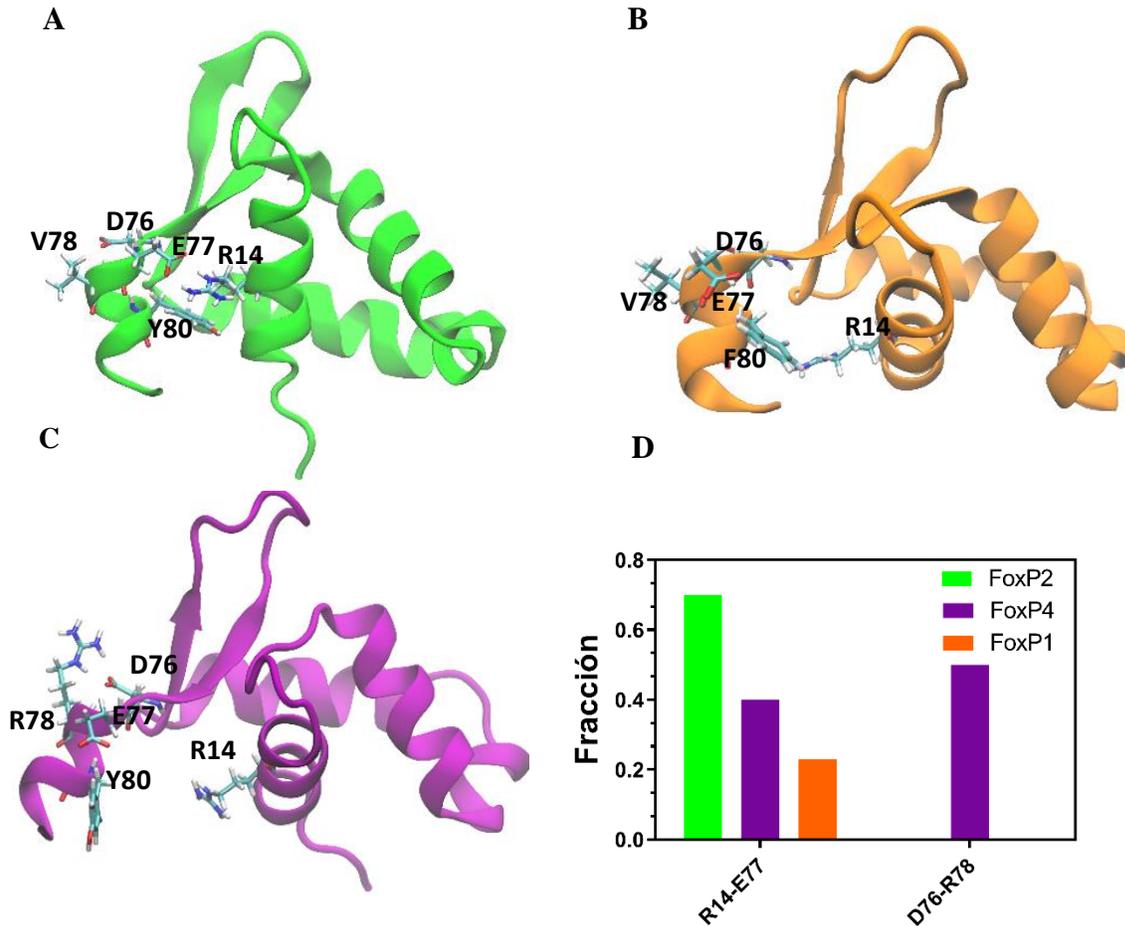
Para estudiar la formación del puente salino que contacta ambas hélices, se cuantificó esta interacción con el *plugin* de VMD *Salt Bridges*, usando una distancia de 3,5 Å entre el centro de masa de los átomos de oxígeno de la cadena lateral de E77 y el centro de masa de los átomos de nitrógeno de la cadena lateral de R14. Los resultados muestran que este puente salino está muy favorecido en FoxP2 (Figura 13 A y D), mientras que está más desfavorecido en FoxP4 y FoxP1, respectivamente. En efecto, la pérdida de este puente salino y la reorientación de Y80 permite la exploración de otras conformaciones en FoxP1 y FoxP4 de esta hélice (Figura 13 B y Figura 13 C), aunque la conformación cerrada sigue siendo predominante en estas simulaciones, pero no en la magnitud de FoxP2.



**Figura 12. Análisis de estabilidad conformacional de los monómeros FoxP.** A. Distribución del radio de giro (Rg) de las simulaciones moleculares de FoxP2 (verde), FoxP1 (naranja) y FoxP4 (morado). Superposición de las estructuras de los centroides obtenidas a partir del análisis de clústeres *k-means*. Se aprecia que los centroides tienen pocas fluctuaciones atómicas, excepto en la región destacada por la flecha negra. La flecha roja destaca el lazo que conecta la hélice H1 y la hebra S1. La flecha negra destaca la hélice H5, que es la región de la proteína con mayor flexibilidad conformacional durante las simulaciones.

**Tabla 6. Abundancia de los *clusters* identificados en las simulaciones de los monómeros FoxP.**

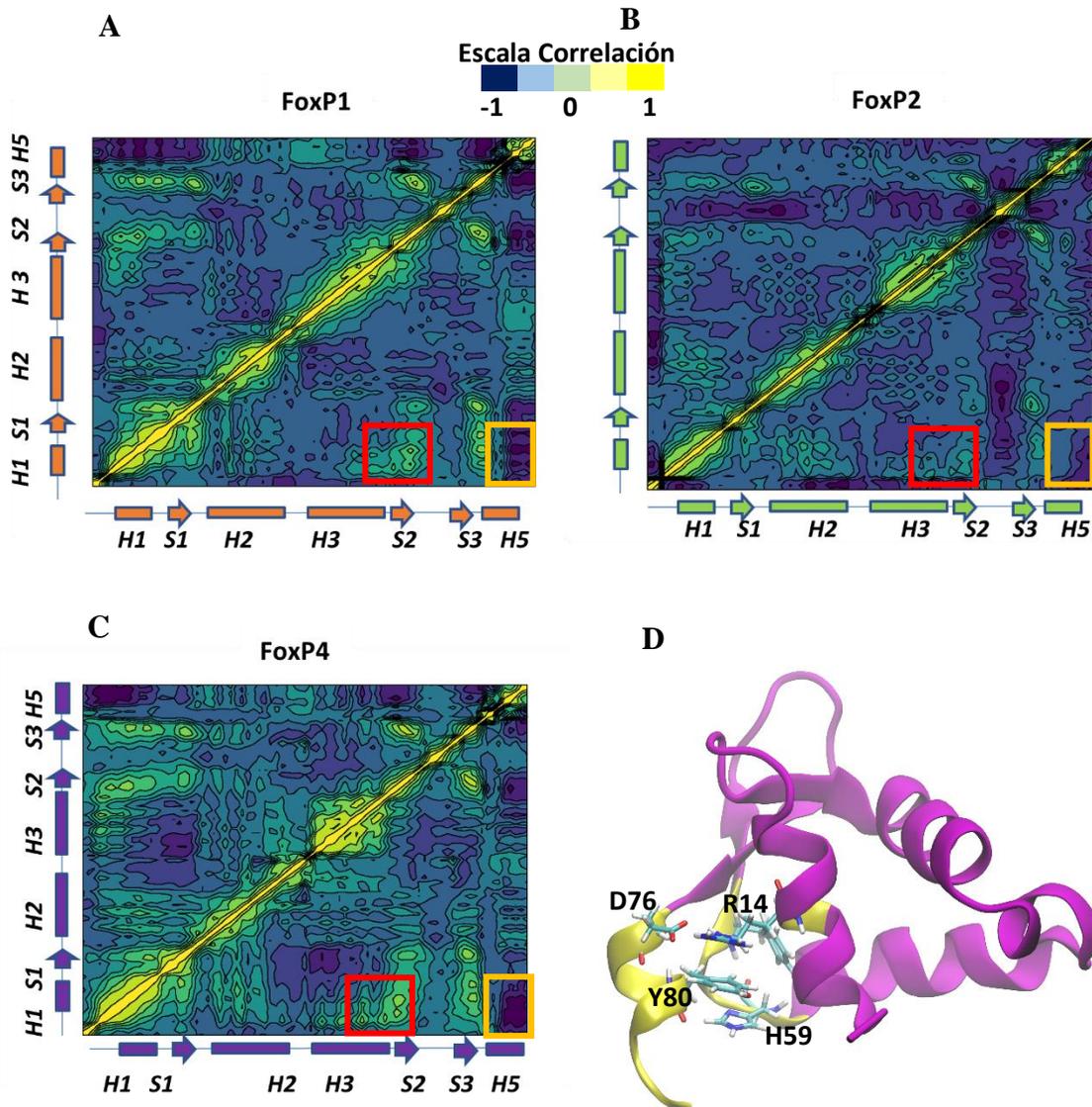
<b>Clúster N°</b>	<b>FoxP2 (%)</b>	<b>FoxP1 (%)</b>	<b>FoxP4 (%)</b>
0	23,2	16,7	21,1
1	13,7	12,3	20,0
2	12,5	11,0	14,3
3	12,2	10,6	10,6
4	11,3	10,5	8,0
5	9,6	10,1	6,7
6	7,3	9,8	6,5
7	7,3	9,2	4,9
8	2,1	6,4	4,1
9	0,1	3,3	3,0



**Figura 13. Interfaz entre las hélices H1-H5 de los monómeros FoxP** **A.** Conformaciones representativas de las estructuras de FoxP2 cerrada, como la resulta en la estructura cristalográfica (PDBid:20A7). **B.** Cuadro representativo de la conformación de FoxP1 donde se pierde la interacción R14-E77. **C.** Cuadro representativo de la conformación de FoxP4 donde se pierde la interacción R14-E77. **D** Cuantificación de los puentes salinos R14-E77 y D76-R78. Las interacciones fueron cuantificadas en la herramienta de VMD *Salt Bridges*, usando como distancia de interacción 3,5 Å entre dador y aceptor.

Los análisis previos de las simulaciones moleculares sugieren que gran parte de las diferencias entre los monómeros FoxP se dan en la interfaz formada entre las hélices H1 y H5. Para buscar vías de comunicación de residuos dentro de la estructura de estas proteínas, se estudió las correlaciones de los movimientos atómicos. En estos análisis, se obtiene como resultado una matriz que muestra la correlación entre el movimiento de cada uno de los residuos de la proteína, cuyos valores van desde -1 (movimiento anti-correlacionado o correlación negativa), 0 (sin correlación) y 1 (movimiento correlacionado o correlación positiva). El movimiento de dos aminoácidos está correlacionado si durante la trayectoria ambos se mueven en la misma dirección, por el contrario, si están correlacionados, pero se mueven en direcciones opuestas se dice que están anti-correlacionados. El análisis de la correlación entre pares o grupos de residuos ayuda a conocer si regiones específicas de la proteína muestran movimientos concertados durante la simulación, dando cuenta de movimientos colectivos y que pueden llegar a ser relevantes a nivel funcional (Ichiye & Karplus, 1991)

Las trayectorias de cada una de las proteínas fueron analizadas y las matrices de correlación resultantes se muestran en la Figura 14. La comparación de ellas muestra que existen regiones en donde hay claras diferencias entre estas proteínas. Una de las más notorias es la que se observa entre la hélice H5 y H1 que, de acuerdo con lo expuesto en las secciones anteriores, representa un movimiento de separación de ambas hélices en FoxP1 y FoxP4 respecto a la conformación cerrada observada para FoxP2. Esa región precisamente se observa con un movimiento anti-correlacionado (azul oscuro) en FoxP1 y FoxP4 mientras que en FoxP2 esa anti-correlación es más leve (azul claro). El análisis específico de los valores de las correlaciones obtenidos entre un residuo representante de la hélice H1 como R14 y parte de los residuos de la hélice H5 se muestran en la Tabla 7 y también se representan en parte en el recuadro naranja de la figura 14. Una diferencia llamativa es la que se aprecia entre R14 e Y80 de FoxP1 (-0,60) y FoxP4 (-0,52) con lo observado en FoxP2, que muestra una correlación cercana a 0 (-0,06), en buen acuerdo con los análisis de la persistencia de este puente salino (R14-E77) en donde Y80 contacta a R14 por medio de un  $\pi$ -*stacking* desde la cara expuesta al solvente.



**Figura 14. Análisis de correlación de los movimientos atómicos de los monómeros FoxP.** Las matrices de correlación de  $C\alpha$  fueron graficadas en representación de superficie en una escala de -1 (azul: movimientos anticorrelacionados o correlación negativa), 0 (verde: ausencia de correlación) y 1 (amarillo: correlacionados o correlación positiva). Los elementos de estructura secundaria correspondientes a cada zona de la matriz están representados en cada matriz. **A.** Matriz de correlación de la simulación molecular del monómero de FoxP1. **B.** Matriz de correlación de la simulación molecular del monómero de FoxP2. **C.** Matriz de correlación de la simulación molecular del monómero de FoxP4. **D.** Estructura del monómero de FoxP4, destacado en amarillo la vía de comunicación entre R14 (Hélice H1) y residuos de H5, y la región desestructurada entre H3 y S3 (Tabla 8). En representación de esferas se destaca cada  $C\alpha$  de los residuos correspondientes. Los valores de correlación para esos residuos se encuentran en las tablas 7 y 8.

**Tabla 7. Correlaciones entre hélice H1 (R14) y carboxilo terminal de hélice H5.**

	F/Y80	Q81	K82	R83	R84
FoxP1-R14	-0,60	-0,58	-0,51	-0,60	-0,54
FoxP4-R14	-0,52	-0,59	-0,57	-0,62	-0,55
FoxP2-R14	-0,06	-0,20	-0,24	-0,32	-0,28

**Tabla 8. Correlaciones entre hélice H1 (R14) y región entre H3-S2.**

	H59	K60	C61	F62	V63
FoxP1-R14	0,39	0,42	0,52	0,64	0,51
FoxP4-R14	0,43	0,48	0,62	0,61	0,50
FoxP2-R14	0,14	0,16	0,27	0,31	0,21

Otro cambio interesante se aprecia en lo que ocurre con las hebras S2 y S3. En esos elementos de estructura secundaria, existen posiciones específicas en donde tanto FoxP1 como FoxP4 exhiben una correlación positiva fuerte (amarillo) y en FoxP2 se aprecian valores que indican una ausencia de correlación (verde). Los valores expresados en la Tabla 8, que representan parte de los observados en el recuadro rojo de la figura 14, indican que la correlación entre R14 de la hélice H1 y los residuos de la región entre la hélice H3 y hebra S2 como F62 y a la región del carboxilo terminal de la hélice H3 como H59 están correlacionados, sugiriendo que el movimiento de ambas regiones es concertado.

Estos resultados, que sirvieron para estudiar vías de comunicación en el estado nativo de los monómeros FoxP, muestran que la hélice H1 se conecta con residuos de la región entre la hélice H3 (F62, C61, H59) y hebra S2. Precisamente, y de acuerdo con los resultados publicados, se ha demostrado que la protonación de H59 en esa región controla la dimerización al afectar principalmente la cinética de asociación de los monómeros (Medina et al., 2019). Además, residuos de esa región como H59, en la hélice H3, aportan con contactos de su cadena principal para formar el sitio de unión a sodio que se observa en la estructura de FoxP4 reportada en esta tesis (PDBid: 6XAT), sugiriendo que este sitio podría también verse afectado por la presencia de iones y que su presencia podría tener un impacto en regiones alejadas del sitio de unión. En efecto, y pese a solo investigar la dinámica conformacional del estado nativo, el hecho de que haya una vía de comunicación entre residuos de las hélices H1 y la región H3-S3 da indicios de las regiones que poseen movimientos concertados en el monómero. Estas regiones podrían ser susceptibles a generar un cambio conformacional que dé lugar a eventos de desplegamiento local o cambios conformacionales más grandes para provocar, por ejemplo, la exposición de residuos para favorecer la asociación por intercambio de segmentos.

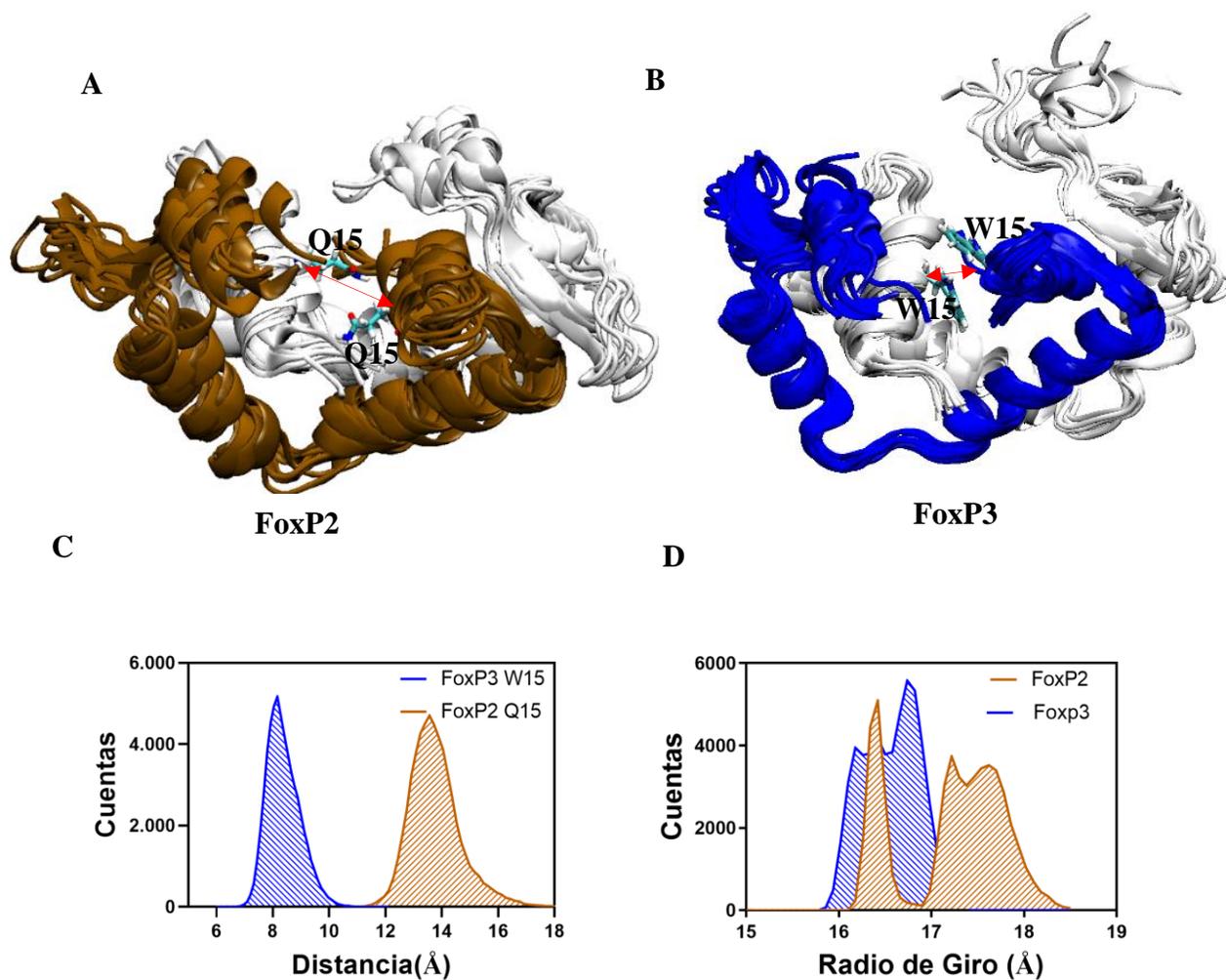
### 3.9. Dinámica conformacional de los dímeros de FoxP2 y FoxP3.

Los datos experimentales sugieren que las principales diferencias de la capacidad de dimerizar en FoxP1 y FoxP4 están dados mayormente en la barrera cinética de asociación. Sin embargo, aunque en menor medida, estas proteínas también exhiben diferencias en la barrera cinética de disociación, la cual es menor para FoxP4 que para FoxP1. Según eso, es muy probable que, en el caso de FoxP2, las diferencias en las barreras cinéticas de asociación y disociación con FoxP1 sea aún más marcadas, ya que esa proteína tiene una constante de disociación al equilibrio del orden milimolar (Perumal et al., 2015). Por otro lado, FoxP3 ha sido caracterizado como un dímero obligado cuyas diferencias en las propiedades de oligomerización serían producto de algunos cambios de secuencias en las hélices H1 y H2 (Bandukwala et al., 2011). Estos elementos son consistentes con trabajos previos que sugieren que en FoxP3 hay una interfaz secundaria adicional a la formada por los contactos inter-cadenas en el dímero con intercambio de segmentos (Bandukwala et al., 2011). En ese contexto, se estudió la dinámica conformacional de los dímeros de la familia FoxP para entender cuáles son los contactos y las propiedades de la estructura a nivel global que diferencian a estas proteínas.

Para eso se utilizó la estructura de los dos dímeros disponibles para estas proteínas: FoxP2 (PDBid 2A07) y FoxP3 (PDBid 3QRF). Las simulaciones moleculares de los dímeros fueron realizadas en triplicado y siguiendo un protocolo similar al de los monómeros. Su fase de producción tuvo una duración de 180 ns contemplando 158 residuos por proteína (79 por cadena polipeptídica).

Las simulaciones moleculares de FoxP2 y Foxp3 muestran a nivel de la estructura global diferencias en su variabilidad conformacional a lo largo de las simulaciones. En esa línea, las trayectorias fueron sometidas a un análisis de clústeres con el algoritmo *k-means* seleccionando 10 clústeres y se identificó las estructuras más representativas (centroides) de cada sistema simulado, la abundancia de cada clúster se presenta en la Tabla 9. La superposición de la estructura de los centroides, las estructuras representativas de cada cluster, muestra que en FoxP2 las hélices H1 y H2 tienen más variabilidad conformacional que el resto de la estructura en esa proteína (Figura 15 A). Por el contrario, en el caso de FoxP3, solo una de las hélices H5 se despliega, pero el resto de la estructura permanece

estable en todos los clústeres (Figura 15 B). Específicamente, respecto a la orientación de las hélices H1, se observa una diferencia en las dos estructuras que es retenida por todos los clústeres de las proteínas: en FoxP2 ambas hélices en el dímero aparecen orientadas paralelamente, mientras que en FoxP3 una hélice se gira respecto a la otra. Al orientarse de esta forma, ambas hélices H1 de FoxP3 favorecen la interacción de W15 y otros residuos hidrofóbicos que son parte de esta interfaz secundaria característica de esta proteína (Bandukwala et al., 2011). Para evaluar las distancias de las hélices H1, se calculó las distancias interatómicas entre W15 en FoxP1 y Q15 en FoxP1 (Figura 15 C). La reorientación de la hélice H1 de FoxP3 respecto a FoxP1 nivel global parece compactar la estructura de FoxP3 respecto a FoxP2, tal como lo indica un análisis de radio de giro de las simulaciones, en donde todas las conformaciones de FoxP3 fluctúan entre un radio de giro de 16 Å a 17 Å aproximadamente (Figura 15 D). Por otra parte, la mayoría de las conformaciones de FoxP2 están sobre esos valores, sugiriendo que FoxP2 explora conformaciones en un estado más abierto que el observado en FoxP3 (Figura 15 D).



**Figura 15. Análisis de las simulaciones moleculares de los dímeros de FoxP2 y FoxP3.** Para evaluar la heterogeneidad conformacional durante las trayectorias estas fueron analizadas con el algoritmo de clusterización *k-means*, seleccionando 10 clústeres de los que se destacan las 10 estructuras centroides correspondientes a cada clúster de FoxP2 (A) y FoxP3 (B). C. Distancia interatómica entre los C $\alpha$  (flecha roja) de W15 en FoxP3 y Q15 en FoxP2. D. Distribución del radio de giro de las simulaciones moleculares de FoxP3 y FoxP2.

**Tabla 9. Abundancia de los clústeres de las simulaciones de los dímeros de FoxP2 y Foxp3.**

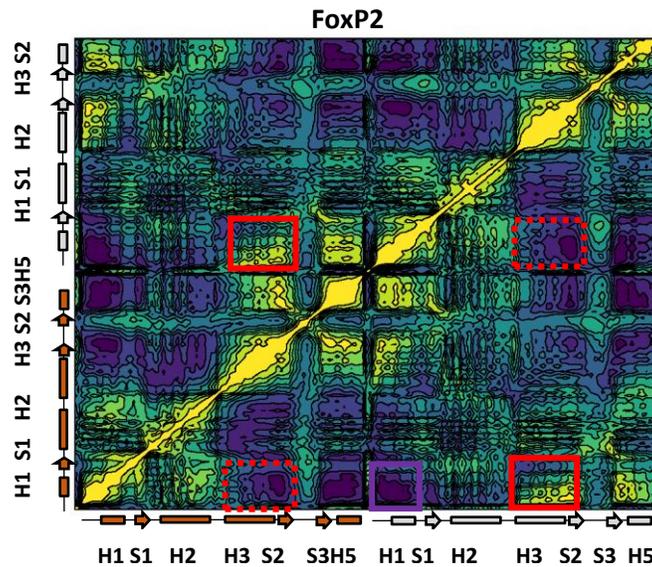
<b>Clúster N°</b>	<b>FoxP2 (%)</b>	<b>FoxP3 (%)</b>
0	29,3	18,1
1	16,6	17,3
2	12,8	12,0
3	12,2	11,8
4	8,2	11,4
5	5,5	8,5
6	5,1	7,7
7	4,8	5,9
8	3,3	4,7
9	2,3	2,5

Para evaluar el efecto de la interfaz adicional en FoxP3 en la dinámica de la proteína se analizaron los movimientos correlacionados para ambos dímeros. Las matrices de correlación se muestran en la Figura 16, en donde se encuentran esquematizados los elementos de estructura secundaria de cada cadena polipeptídica al costado de cada matriz. Las matrices muestran diferencias claras entre ambas proteínas, muchas regiones de FoxP2 tiene una matriz con movimientos fuertemente anti-correlacionados (azul oscuro), mientras que FoxP3 presenta una matriz con correlaciones más bajas (azul o verde claro). Los movimientos correlacionados (amarillo) se dan mayormente en la diagonal de la matriz de ambas proteínas y en algunas zonas como muy localizadas. La interfaz que forman las hélices H1 es una de las partes de la estructura que provee de una superficie adicional de contacto entre ambas hélices y también con otros elementos como la región de la bisagra en la hélice H2. Al observar los movimientos correlacionados en ambas hélices H1 de la interfaz (cuadro morado), se aprecia que en FoxP3 hay correlaciones muy bajas y cercanas a cero (azul claro o verde claro) y en FoxP2, esa región tiene movimientos anti correlacionados (azul oscuro). Parte de las correlaciones calculadas para la hélice H1 se muestran en la Tabla 10, en donde se escogió a dos residuos de la H1 de la cadena A (R14 y W15) y a un grupo de residuos de la hélice H1 de la cadena B (5-10), en donde se puede corroborar lo observado en la matriz de correlaciones. El antecedente de que FoxP3 tenga un bajo grado de correlación en esta región es sugerente de la poca variabilidad conformacional (Figura 15) durante las trayectorias, en cambio FoxP2 tiene una alta anti-correlación (azul oscuro), indicando que ambas hélices se mueven en dirección opuesta, por lo tanto, evitando la formación de la interfaz. Esto es coincidente también con los análisis que indican que FoxP2 es una proteína que es menos compacta que FoxP3 (Figura 15.D). Además, el mismo grupo de residuos que estaba correlacionado en los monómeros de FoxP4 y FoxP2 (recuadro rojo en sus respectivas matrices de correlación, Figura 14) aparece con correlaciones positivas en FoxP2 y también en FoxP3, pero en el caso de la última en menor medida (Figura 16, recuadro rojo y Tabla 11). En ese caso, es importante destacar que en cada subunidad del dímero los contactos que estabilizan esas regiones (H1-S1 con H3-S2) provienen de diferentes cadenas polipeptídicas, en cambio en el monómero provienen de la misma cadena polipeptídica. Por lo mismo, al observar en los dímeros la correlación de los movimientos atómicos de esas regiones en la misma cadena polipeptídica (Figura 16, cuadrado punteado rojo), H1-S1 pertenece a una

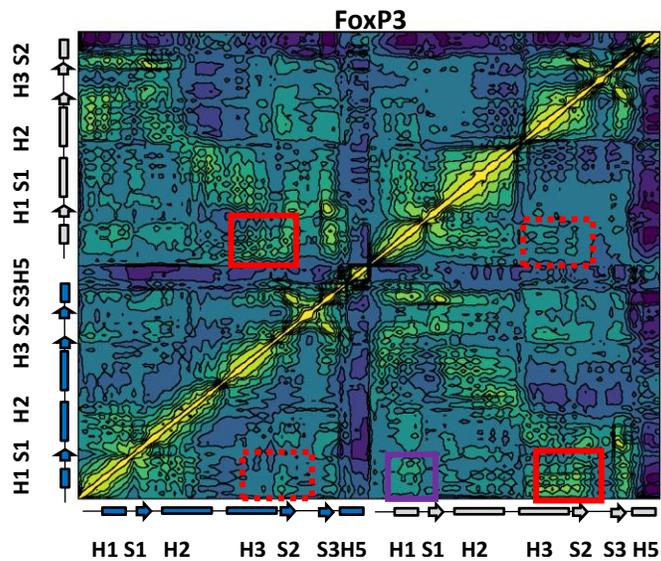
subunidad en el dímero y H3-S3 a otra, ambas muestran una alta anti-correlación (azul), sugiriendo que las dos subunidades del dímero están moviéndose en direcciones opuestas en FoxP2, y, por lo tanto, alejándose. Por el contrario, esa zona en FoxP3 aparece con una anticorrelación muy baja (azul claro).

Estos análisis sugieren que el dímero de FoxP2, el que está muy desfavorecido por su elevada constante de disociación, tiene una alta variabilidad conformacional en la subfamilia FoxP. Por el contrario, FoxP3, el dímero más estable tiene una baja variabilidad conformacional durante las simulaciones. Si bien, en este trabajo de tesis se ha determinado que la asociación de los monómeros es la mayor barrera energética que deben superar estas proteínas para estabilizar el dímero con intercambio de segmentos, la baja variabilidad conformacional de FoxP3 respecto a lo observado en FoxP2, que es más flexible, es un indicativo que en el caso de la primera los contactos y residuos que las diferencian han establecido redes de residuos que determinan las propiedades del estado nativo de ambas.

A



B



**Figura 16. Matrices de correlación de las simulaciones moleculares de los dímeros de FoxP2 y FoxP3.** Las matrices de correlación de C $\alpha$  fueron graficadas en representación de superficie en una escala de -1 (azul: movimientos anti correlacionados), 0 (verde: ausencia de correlación) y 1 (amarillo: correlación). Los elementos de estructura secundaria correspondientes a cada zona de la matriz están representados en cada matriz

**Tabla 10 Correlaciones entre residuos inter-subunidad de las hélices H1.**

	5 (A)	6 (A)	7 (A)	8 (A)	9 (A)	10(A)
FoxP2 R14 (B)	-0,81	-0,84	-0,83	-0,83	-0,86	-0,77
FoxP2 W15 (B)	-0,79	-0,80	-0,76	-0,76	-0,79	-0,67
FoxP3 R14 (B)	0,23	0,16	0,24	0,25	0,24	0,30
FoxP3 W15 (B)	0,21	0,167	0,26	0,25	0,26	0,37

**Tabla 11. Correlaciones entre H1 (R14) y la región entre H3-S2**

	54	55	56	57	58	59
FoxP2(14A)	0,68	0,70	0,82	0,78	0,67	0,44
FoxP2(14B)	0,72	0,73	0,83	0,80	0,71	0,51
FoxP3(14A)	0,58	0,52	0,62	0,65	0,52	0,39
FoxP3(14B)	0,39	0,39	0,52	0,62	0,51	0,38

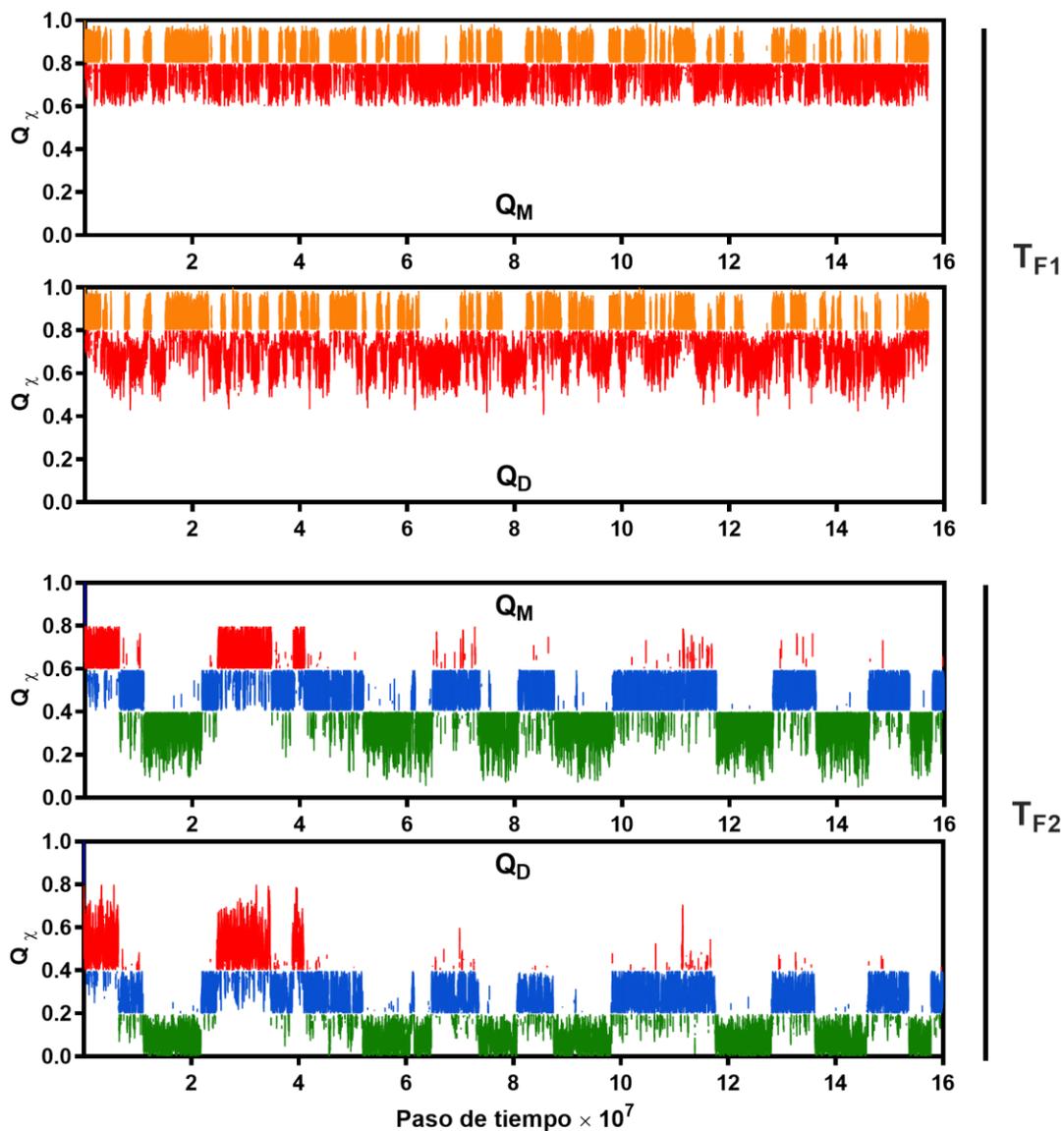
### 3.10. Simulaciones moleculares de plegamiento del dímero de FoxP1.

Las simulaciones moleculares convencionales mostraron que el dímero de FoxP3 es una especie altamente estable y el dímero de FoxP2 uno mucho más flexible conformacionalmente. Sin embargo, estudios recientemente publicados exploraron en profundidad la heterogeneidad conformacional del dímero de FoxP1 usando FRET de molécula única y simulaciones moleculares. Los estudios de FRET de molécula única se realizaron marcando distintas regiones de los dímeros de FoxP1 con sondas fluorescentes, obteniendo así las distancias observadas para diferentes pares de residuos en el dímero en solución. Esos experimentos sugirieron que el dímero de FoxP1 es conformacionalmente heterogéneo, encontrándose no solo en la conformación canónica del dímero con intercambio de segmentos observada en el dímero de FoxP2 y FoxP3, sino que también tiene conformaciones más extendidas que contienen regiones desplegadas de la proteína. Para proponer una explicación a nivel atómico de esos resultados experimentales, en esta tesis se realizaron simulaciones moleculares de plegamiento del dímero de FoxP1 utilizando modelos nativo-céntricos cuyos resultados fueron publicados (Medina et al., 2020),

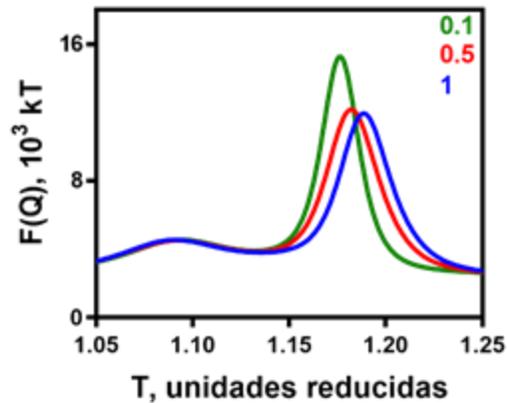
Las simulaciones de plegamiento proteico fueron realizadas usando una función de energía que se basa en los contactos nativos observados en la estructura de la proteína, los cuales son calculados a partir de un criterio de distancia máxima átomo-átomo entre residuos cuya separación en la secuencia primera es, al menos, tres residuos. La función de energía corresponde a un potencial de Lennard-Jones, el cual considera a todas las interacciones nativas átomo-átomo como atractivas cuyo mínimo energético se encuentra a la distancia de la estructura inicial, mientras que el resto de las interacciones que no están presentes en la estructura inicial son evaluadas con el componente repulsivo de la ecuación de Lennard-Jones. Estos modelos además incluyen una representación simplificada de los átomos (*all-atom*) de la estructura de las proteínas, que no presenta hidrógenos y en la cual la masa de todos los átomos es unitaria. Todas estas simplificaciones de la función de energía permiten estudiar eventos moleculares que ocurren en escalas temporales superior a las simulaciones convencionales como vías de plegamiento de proteínas (Noel et al., 2010) (Ramírez-Sarmiento et al., 2015).

Para realizar estas simulaciones, la estructura del dímero de FoxP1 se obtuvo mediante modelamiento por homología, usando como molde la estructura dimérica de FoxP2 (PDBid: 2A07). Con la estructura modelada de FoxP1 se realizaron simulaciones moleculares a temperaturas bajo y sobre la temperatura de plegamiento ( $T_F$ ) observándose varias transiciones de plegamiento (Figura 17). Se simuló el efecto de la concentración de proteína efectiva usando restricciones armónicas sobre los centros de masa de los monómeros a diferentes a diferentes valores de fuerza de ese potencial (Figura 18). Para seguir las simulaciones, como coordenada de reacción se utilizó los contactos nativos del monómero (intra-cadena,  $Q_M$ ) y dímero (inter-cadena,  $Q_D$ ), por su habilidad para discriminar entre los diferentes pozos energéticos observados durante el plegamiento de la proteína.

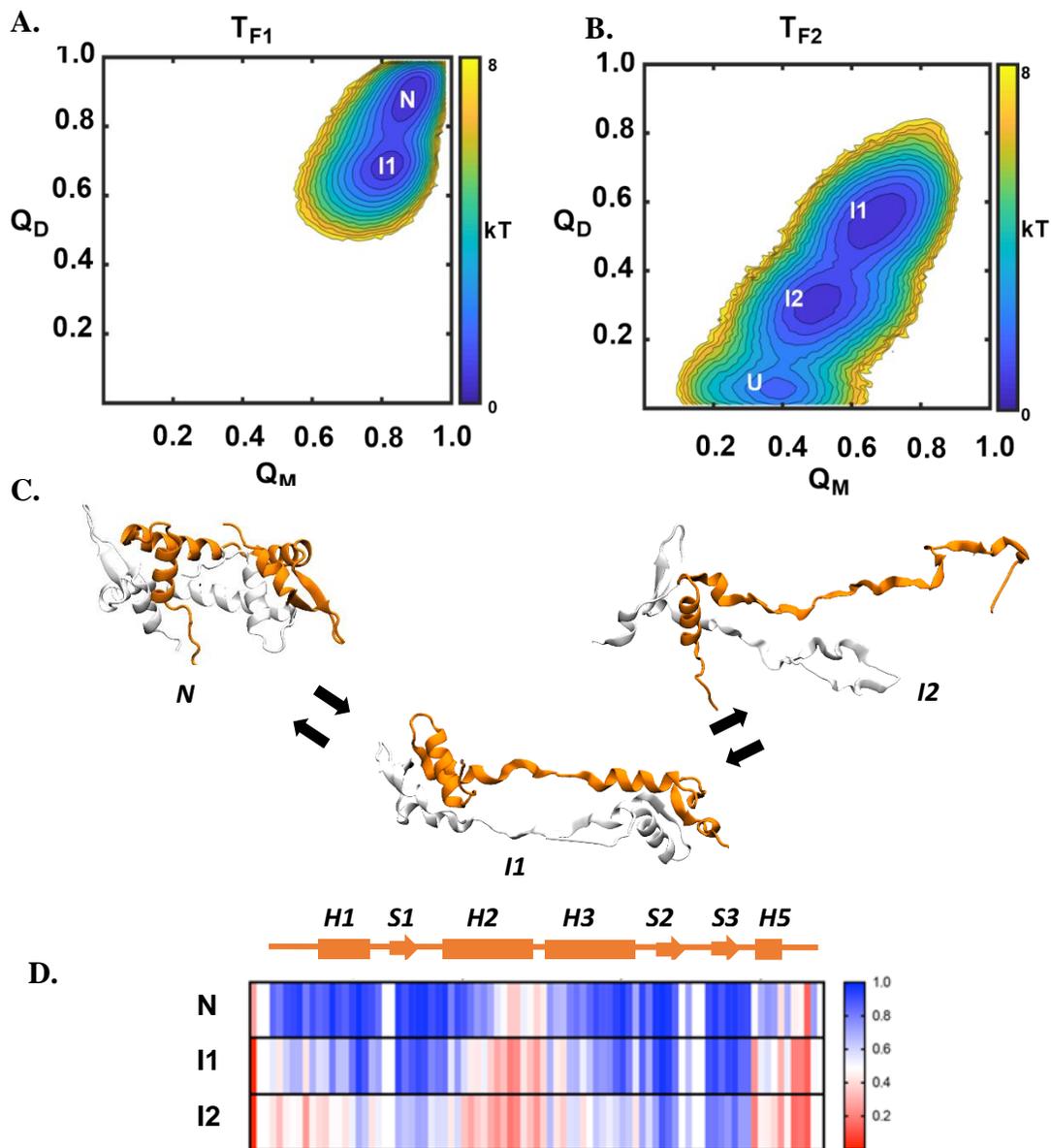
El análisis de los cambios en calor específico de las simulaciones mostró dos temperaturas en donde ocurren transiciones de plegamiento,  $T_{F1}$  y  $T_{F2}$  (Fig 18). Con la información estructural y energética de las simulaciones moleculares realizadas a temperaturas diferentes se utilizó el método de análisis de histogramas (WHAM) (Kumar et al., 1992b), usando como coordenadas de reacción los contactos nativos ( $Q_D$  y  $Q_M$ ) lo que permitió generar el paisaje energético de las simulaciones a las temperaturas  $T_{F1}$  y  $T_{F2}$  (Figura 19 A y 19 B). Los paisajes energéticos a ambas temperaturas muestran pozos de mínima energía que representan ensamblajes conformacionales diferentes, solo separados por pequeñas barreras energéticas (1-2 kt, Figura 19 A y Figura 19 B), sugiriendo que el tránsito entre desde un ensamblaje a otro es muy frecuente. Consistentemente, los análisis de las conformaciones observadas durante las trayectorias muestran múltiples transiciones entre cada estado (Fig. 17). En el caso de  $T_{F1}$ , se observa un pozo energético donde  $Q_D$  y  $Q_M$  representan una fracción cercana a 1, ese pozo energético corresponde al ensamblaje conformacional del dímero con intercambio de segmentos (DS). Ese mínimo energético está conectado con otro en donde  $Q_D$  es  $\sim 0,8$  y  $Q_M \sim 0,6$  (I1), el que representa un ensamblaje conformacional que pierde parte de los contactos de la región bisagra, resultando en un dímero extendido (Figura 19 C). En el caso de  $T_{F2}$ , se observan dos pozos energéticos previos a la observación del estado desplegado ( $Q_D \sim 0,1$ ,  $Q_M \sim 0,4$ ). Uno de ellos corresponde al ensamblaje conformacional I1, el cual está conectado a un segundo ensamblaje conformacional (I2) en donde los contactos nativos  $Q_D$  y  $Q_M$  se reducen casi a  $\sim 0,4$ , lo que resulta en una subunidad desplegada mientras que la otra se mantiene plegada (Figura 19 C).



**Figura 17.** Cambios de la fracción de los contactos nativos monoméricos (QM) y diméricos (QD) en las temperaturas de plegamiento  $T_{F1}$  y  $T_{F2}$ . En  $T_{F1}$  se aprecian múltiples transiciones entre el dímero nativo (DS, naranja), ensamble I1 (rojo). Lo mismo ocurre en  $T_{F2}$  entre el ensamble I1, ensamble I2 (azul) y el estado desplegado U (verde).



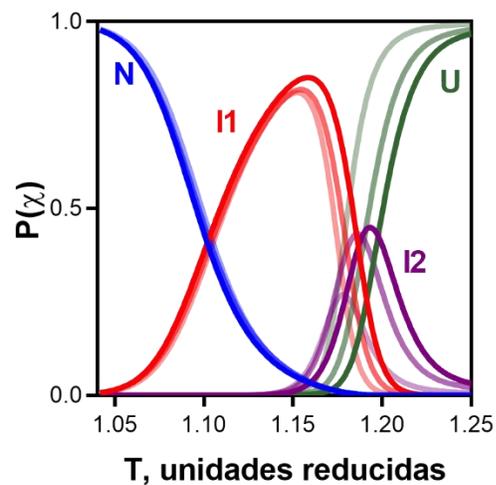
**Figura 18. Cambios en la capacidad calórica en las transiciones de plegamientos a diferentes concentraciones efectivas de proteína.** El cambio en la capacidad calórica a las diferentes concentraciones se calculó usando WHAM. La concentración de proteína fue variada usando una restricción harmónica ( $k$ ) entre los centros de masas de cada monómero a una distancia observada en el dímero con intercambio de segmentos (DS), configurando el valor de esta constante a 0,1 (verde), 0,5 (rojo), y 1 (azul)  $\epsilon \cdot \text{nm}^{-2}$ . Se observa un pequeño pico a  $T=1,09$  (TF1), el que como muestra la figura, no se ve afectado por la concentración de proteína. El segundo pico se observa a  $T=1,18$  (TF) a temperaturas elevadas, pero decrece en las temperaturas bajas cuando la concentración efectiva disminuye por medio de la disminución de la constante de fuerza  $k$ .



**Figura 19. Simulaciones moleculares del plegamiento del dímero de FoxP1.** Paisaje energético de plegamiento del dímero de FoxP1 a diferentes temperaturas,  $T_{F1}$  y  $T_{F2}$  corresponden a los dos picos de cambios en el calor específico en función de la temperatura. **A.** El paisaje energético de  $T=T_{F1}$  muestra dos pozos energéticos, dímero nativo (N1) e Intermediario 1 (I1). **B.** Por el contrario, el paisaje energético de  $T=T_{F2}$  muestra tres pozos, los que corresponden al intermediario 1 (I1), intermediario 2 (I2) y al estado desplegado (U). **C.** Estructuras representativas de los ensambles conformacionales I1, I2 y N. **D.** Probabilidad de contacto de los tres ensambles conformacionales observados durante de la simulación  $T=T_{F2}$ . La escala representa ausencia en rojo (0) y presencia en azul (1) de los contactos nativos.

En el estado U se reducen casi la totalidad los contactos nativos  $Q_D$  y se retiene cerca del 0,4 de los contactos intramoleculares o  $Q_M$ , esto último debido a contactos locales que se retienen en este estado de manera transiente. Interesantemente, el estudio del efecto de la concentración de proteína en las fracciones de cada intermediario (Figura 20) muestra que la concentración incrementa la fracción de los ensamblados intermediarios I1 e I2 y desplaza el valor de  $T_{F2}$  hacia temperatura superiores sin afectar  $T_{F1}$  (Figura 18 y 20). Esos resultados sugieren que el ensamblado I2 evita la transición directa de I1 a un estado desplegado cuando se incrementa la concentración de proteína. Considerando los antecedentes expuestos se considera que el mecanismo de plegamiento observado durante las simulaciones del dímero de FoxP1 es:  $I2 \rightleftharpoons I1 \rightleftharpoons DS$ .

Para caracterizar cómo varían los contactos que estabilizan a las estructuras de cada ensamblado conformacional durante las simulaciones de  $T_{F1}$  y  $T_{F2}$  se aislaron sus estructuras de las simulaciones moleculares y se calculó la probabilidad de contacto de cada estado (Figura 19 D). Una fracción cercana a 1 indica la presencia de la totalidad de los contactos intra-subunidad observados en la estructura del modelo inicial (dímero nativo) y un valor de 0 representa la pérdida total de los contactos. Los análisis mostraron que en DS la hélice H5 y parte de la región del carboxilo terminal de la hélice H2 están parcialmente desestabilizados. En efecto, en el ensamblado conformacional I1 esas regiones se desestabilizan casi en su totalidad y se suman algunos contactos de la región amino terminal de la hélice H2. Además, en ese intermediario también se pierden parte de los contactos de la región amino terminal de la hélice H1. Interesantemente, los contactos que comienzan a perderse en DS y luego en I1, reafirman lo observado en las simulaciones moleculares convencionales (Figura 17), que sugerían que la hélice H5 en los dímeros y también monómeros FoxP es altamente flexible, siendo esta la primera hélice en perder su conformación nativa. Finalmente, en el intermediario I2 se terminan de desestabilizar los contactos que conforman la hélice H1, permaneciendo solo una baja fracción de ellos. Solo los contactos de las hebras S2 y S3 se mantienen, siendo éstos los últimos elementos de estructura secundaria observados antes de que la proteína se despliegue.



**Figura 20. Cambios en las fracciones de las poblaciones de los ensambles conformacionales (N, I1, I2, U) en función de la temperatura y concentración de proteína.** Los cambios en la fuerza de la restricción armónica para mantener los monómeros de la proteína cercanos son representados en un degradado de colores entre 0,01, 0,1, 0,5 y 1  $\epsilon \cdot \text{nm}^{-2}$  para cada uno de los ensambles observados durante las simulaciones I1 (rojo), I2 (morado), I1 (rojo) y N (azul).

Los resultados de FRET de molécula única expuestos en el trabajo de Medina y col. (Medina et al., 2020), que incluyen las simulaciones de plegamiento explicadas en la sección anterior, muestran que el dímero de FoxP1 es una proteína altamente flexible y que en solución visita diferentes estados conformacionales, esos estados son posibles producto de eventos de desplegamiento local de la proteína. La presencia de zonas desplegadas en la estructura de FoxP1 incluye grados de libertad de movimiento que le da una flexibilidad conformacional mayor a la del estado nativo y que permitiría a estas proteínas explorar conformaciones que favorezcan la formación del dímero o la unión a otras macromoléculas.

### 3.11. Análisis filogenético de la familia Fox humana.

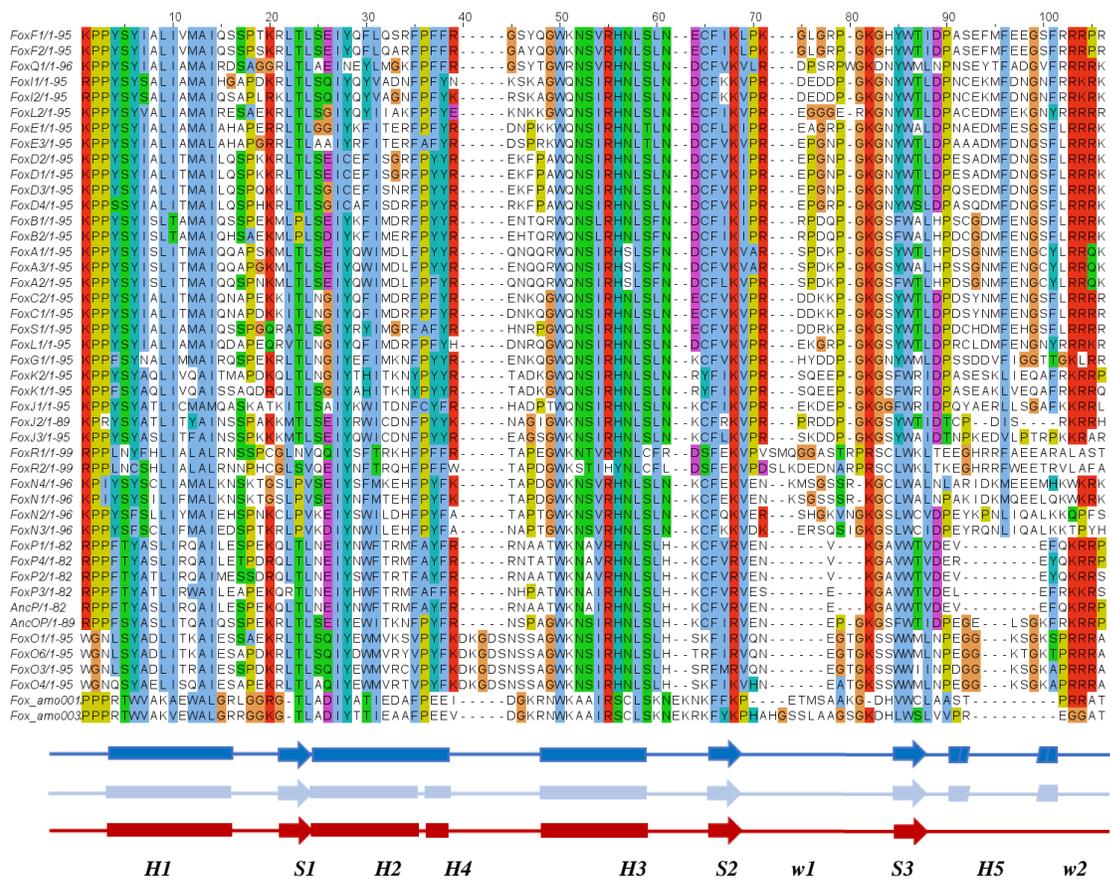
Los cambios de secuencias han modelado a las proteínas FoxP y les han permitido tener propiedades funcionales y de oligomerización diferentes entre sí. Varios de los cambios que explican esas diferencias pudieron haber surgido previo al cambio de prolina por alanina que ha sido caracterizada como la responsable del origen del intercambio de segmentos en la familia Fox. En ese aspecto, es interesante considerar la historia evolutiva de esta familia para contextualizar cómo emergió el intercambio de segmentos en una familia de factores de transcripción que contiene proteínas mayormente monoméricas.

Para entender cómo los cambios en secuencia dentro de la familia Fox permitieron el origen del intercambio de segmentos en esta familia, se construyó un árbol filogenético. Para ello, se descargaron las secuencias de 42 miembros de la familia de la base de datos UNIPROT y fueron alineadas usando el programa Prank (Löytynoja, 2014). En la Figura 21 se muestra el alineamiento resultante, en el cual se observa que varias regiones, en particular en la Hélice H3 de unión a ADN y en la mitad amino terminal de la hélice H1, hay un alto grado de conservación de secuencias en toda la familia Fox. La mayor variabilidad de secuencia se observa en regiones desestructuradas, entre la mitad hacia el carboxilo terminal de la hélice H1, la hélice H2/H4 y en la hélice H5.

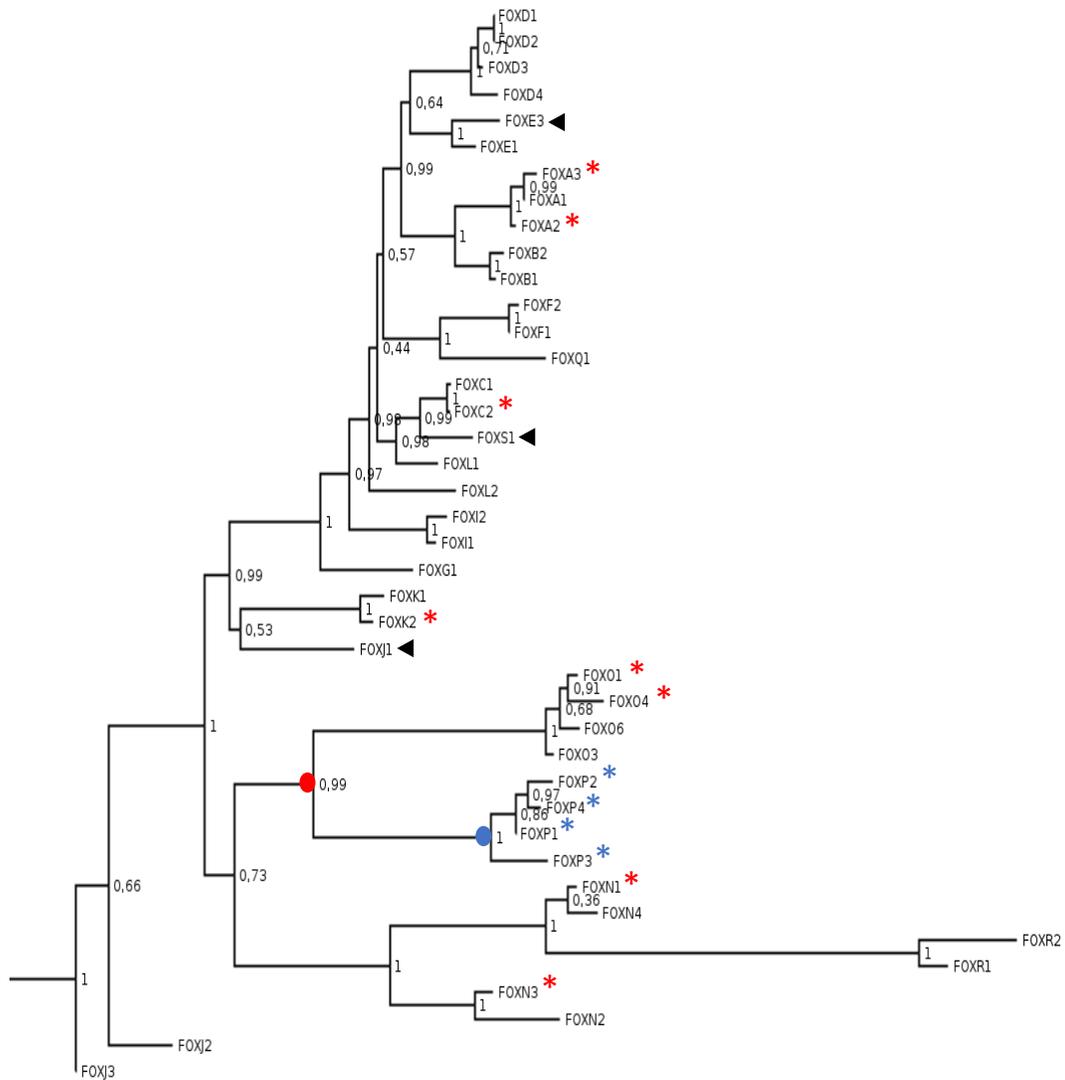
Usando el alineamiento y un modelo de evolución de secuencias se construyó el árbol filogenético de la familia Fox para las proteínas humanas. La topología del árbol resultante muestra varias de las subfamilias correctamente agrupadas respecto a su clasificación (Figura 22). Los miembros con estructura monomérica conocida se distribuyen en gran parte del árbol filogenético de la familia Fox, como se resalta en el árbol filogenético con asteriscos rojos (Figura 22). Es importante notar que existen otros miembros de la familia Fox, aparte de la subfamilia FoxP, que contienen alanina o cisteína en la posición del residuo de la región bisagra que define la oligomerización de FoxP (A39), como por ejemplo FoxJ1, FoxE3, FoxS1 todos ellos destacados con una flecha negra en el árbol filogenético de la familia (Figura 22). No obstante, estos miembros aparecen filogenéticamente muy distantes a la subfamilia FoxP, indicando que la bisagra habría sido aparecido mutada varias veces durante la evolución de esta familia. Lo último podría indicar que esas proteínas pueden oligomerizar por intercambio de segmentos, sin embargo, en una proteína monomérica como FoxK1, el

reemplazo de P39A en la región bisagra no resultó en la formación de dímeros en experimentos usando una concentración de 2mM. (Chu et al., 2011).

El árbol filogenético muestra a todos los miembros de la subfamilia FoxP divergiendo a partir de un ancestro en común (AncP). A partir de AncP se originan dos clados: uno que contiene exclusivamente a FoxP3 y el otro del que emergen FoxP1, FoxP2 y FoxP4. La divergencia de FoxP3 en un clado aislado al resto de las proteínas de la subfamilia es consistente con sus propiedades de oligomerización descritas previamente en este trabajo, que indican que esta proteína es un dímero obligado que exhibe la mayor estabilidad termodinámica entre los miembros de la subfamilia. En el otro clado, se observa la proteína más cercana a AncP, FoxP1, la que tiene una constante de disociación muy baja respecto a FoxP4 y FoxP2, siendo FoxP2 la más distante filogenéticamente al ancestro y la con la mayor constante de disociación, y, por lo tanto, la más propensa a encontrarse como monómero.



**Figura 21. Alineamiento de secuencias de proteínas de la familia Fox de humanos.** Las secuencias de 42 proteínas Fox humanas fueron descargadas de la base de datos UNIPROT y alineadas usando PRANK. Se incluyeron dos secuencias utilizadas como grupo externo en el alineamiento Fox\_amo001 y Fox\_amo003. Además, se incluyeron las secuencias de ambos ancestros (AncP Y AncOP). El alineamiento fue visualizado en Jalview y coloreado según el esquema de colores de ClustalX, en donde cada color representa la conservación de aminoácidos categorizados según su tipo.



**Figura 22. Filogenia de la familia Fox humana.** Usando el alineamiento de la familia se creó el árbol filogenético de 42 miembros de la familia Fox humana usando como grupo externo las secuencias fox\_Amo001 y fox\_amo003, el grupo externo no se incluye para simplificar la visualización del árbol. El círculo rojo representa al ancestro de las subfamilias FoxO y FoxP (AncOP) y el azul al último ancestro en común de la familia FoxP (AncP). Los asteriscos rojos muestran aquellas estructuras con estructuras disponibles en el PDB y que han sido reportadas únicamente como monómeros. En asterisco azul se destacan los cuatro miembros de la familia Fox que han sido caracterizados como dímeros con intercambio de segmentos.

### 3.12. Reconstrucción de Secuencias Ancestrales.

El análisis del árbol filogenético muestra que las proteínas FoxP han divergido desde AncP en dos ramas evolutivas, cuyas proteínas actuales presentan parámetros termodinámicos y cinéticos de oligomerización muy diferentes y, además, que el último ancestro en común entre las proteínas monoméricas y la subfamilia FoxP es el que comparte esta última con la subfamilia FoxO (AncOP). Ambos nodos ancestrales son claves para comprender la historia evolutiva de estas proteínas, tanto para la emergencia del intercambio de segmentos en la familia como para entender las marcadas diferencias en las propiedades de oligomerización de la subfamilia FoxP. Por esta razón, se determinaron las secuencias más probables de ambos ancestros mediante la aproximación de máxima verosimilitud. Para esto se empleó el árbol filogenético (Figura 22) y el alineamiento de la familia Fox (Figura 21) en el programa PAML (Yang, 2007). Para estimar la secuencia ancestral, se calcula la probabilidad de observar cada uno de los 20 aminoácidos posibles en cada posición del alineamiento de secuencias, de esta forma la secuencia final considerada como la secuencia ancestral más probable es aquella que reúne a los aminoácidos con la probabilidad más alta para cada posición. En la figura 23 se muestra las tres probabilidades más altas entre los 20 aminoácidos posibles para cada posición de las secuencias ancestrales calculadas de AncP y AncOP.

La reconstrucción de la proteína AncOP resultó en una secuencia de 87 aminoácidos (Figura 21), y la estimación de la secuencia más probable contiene varias regiones en donde para cada posición no hay una diferencia importante entre la primera y segunda mayor probabilidad de observar un determinado aminoácido (Figura 23 A), lo cual implica que no es posible definir una secuencia única para este nodo. Algunas regiones, como las hélices H2 y H5, tienen posiciones de la secuencia incluso con una primera probabilidad menor a 0,4, resultando la secuencia más probable para AncOP con una probabilidad promedio de 0,77. Por el contrario, la reconstrucción de la secuencia para AncP tiene probabilidades mucho más altas (Figura 23 B), con todas las posiciones del alineamiento presentando un residuo con probabilidad mayor a 0,5, resultando así en una reconstrucción de una secuencia más probable con una probabilidad promedio de 0,94.

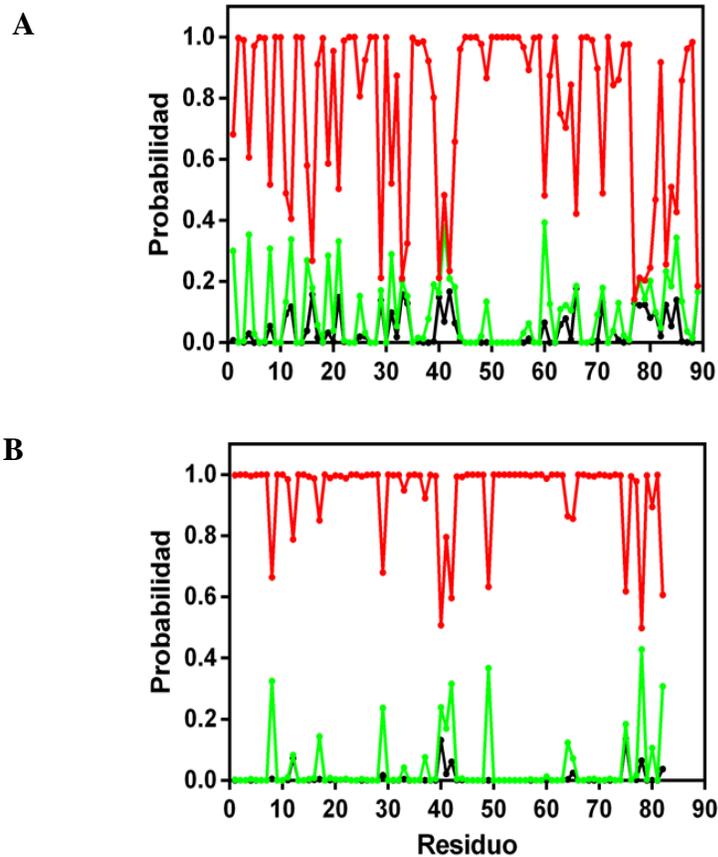
La secuencia más probable calculada para AncP tuvo un mejor soporte que la calculada para AncOP y no se observa ninguna inserción respecto a los miembros actuales de la subfamilia FoxP, por lo que el ancestro retuvo el largo característico de las proteínas FoxP usadas en este trabajo (82 residuos). AncP retuvo la alanina característica de todas las proteínas FoxP en la región bisagra, lo que sugiere que también tenga capacidad de oligomerizar por intercambio de segmentos. Interesantemente, AncP tiene solo dos cambios respecto a la secuencia de FoxP1: un reemplazo en la hélice H3 de unión al ADN en la posición 52 de un residuo de valina por isoleucina y otro en la aleta W1 en la posición 68 de valina por ácido glutámico. El primer cambio tiene una probabilidad de 0,98 y el segundo de 0,68, por lo que ambos están estadísticamente bien soportados. Estas dos sustituciones respecto a la secuencia de FoxP1 son aminoácidos observados en FoxP3 y no en FoxP1, FoxP4 o FoxP2, que muestran en ambas posiciones una valina (Figura 21), lo que augura que este ancestro pueda tener propiedades de oligomerización parecidas a las de FoxP3, si es que ambos son relevantes para la dimerización.

El análisis de la estructura de FoxP3 muestra a I52 en la hélice H3 orientándose hacia el interior de la proteína, formando interacciones con residuos de la misma cadena polipeptídica o con otros provenientes de la hélice H1 de la cadena polipeptídica adyacente en el dímero, formando un núcleo hidrofóbico que podría ser importante para la estabilidad de la proteína. En el caso de E69, corresponde a una posición expuesta a solvente y que ha sido descrita como importante para la interacción con otras proteínas en FoxP3 (Bandukwala et al., 2011), por lo que parecería no tener mayor trascendencia para la dimerización. En efecto, V69E en FoxP1 fue una de las mutaciones propuestas en base a la secuencia descritas en secciones previas de esta investigación (Figura 11), resultando en un leve aumento en la  $K_D$  respecto a FoxP1. Pese al modesto efecto de la mutante V69E en FoxP1, no se puede descartar el efecto de la presencia de ambas mutaciones de forma simultánea como en la secuencia de AncP, especialmente considerando que ambos residuos son observados en FoxP3, una proteína con un dímero mucho más estable.

Si bien los resultados muestran que la reconstrucción de la secuencia más probable para AncP es mucho más robusta que la hecha para AncOP, el análisis de las posiciones de la secuencia de AncOP que tienen un buen soporte permite entender algunos cambios que

debieron ocurrir en la historia evolutiva de esta familia desde ese ancestro a los miembros actuales. La secuencia más probable para AncOP posee el residuo de prolina característica de las proteínas monoméricas en la región bisagra con una probabilidad de 0,98, sugiriendo que este ancestro debiera ser monomérico (Figura 21). Adicionalmente, esta proteína posee varias inserciones respecto a los miembros de la familia FoxP, específicamente entre la aleta 1 (W1) y la hélice H5 en el extremo carboxilo terminal de la proteína. Dentro de las más destacables se puede mencionar la observada en la aleta 1, específicamente AncOP tiene una inserción prolina-glicina (PG) respecto a la secuencia FoxP, la que se observa en gran parte del resto de la familia Fox (Figura 13). Mientras que en el motivo PG, la presencia de una G está soportada por una probabilidad de 1, la de prolina solo por 0,42. Aunque baja, la siguiente probabilidad en esa posición es para una glicina con un 0,19, por lo que la preferencia es clara por prolina. Otra diferencia importante respecto a la subfamilia FoxP es la presencia de una prolina en la región amino terminal de lo que correspondería a la hélice H5 en esas proteínas. La presencia de prolina está soportada con una probabilidad de 0,98 en la secuencia de AncOP, mientras que en las proteínas FoxP se conserva estrictamente un ácido glutámico en esta posición. La presencia de esta prolina en las proteínas monoméricas y en AncOP podría explicar que esta alfa hélice sea una zona desestructurada o desestabilizada, como se observa en varios de los miembros actuales de la familia Fox como FoxM1 (PDBid: 3G73) (Littler et al., 2010) o FoxC2 (PDBid: 6AKP) (X. Chen et al., 2019).

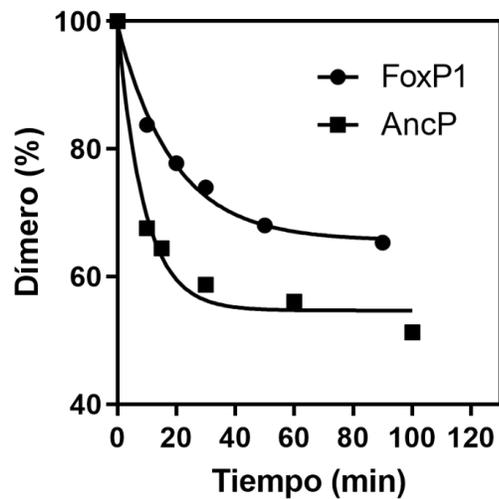
Estos antecedentes sugieren que el intercambio de segmentos emergió en esta familia posterior a la aparición del ancestro monomérico AncOP, por medio del reemplazo de alanina por prolina en la región bisagra observada en AncP. En adición a esa mutación, otros cambios ocurrieron durante la evolución que originaron a los miembros de la subfamilia FoxP y el AncP, destacando entre ellos la delección de los residuos anteriormente mencionados en W1 y H5. Por lo que, en el contexto de la historia evolutiva de la familia Fox, no fue solo necesaria la mutación de alanina por prolina en la región bisagra, sino que también esas delecciones ocurrieron previo a que el dímero con intercambio de segmentos emergiera en esta familia de factores de transcripción.



**Figura 23 Probabilidades de las secuencias de AncOP y AncP.** Las secuencias más probables calculadas para cada ancestro (color rojo) se alinearon con el resto de la familia Fox en la Figura 21. Los residuos con las tres probabilidades más altas entre los 20 aminoácidos posibles para cada posición de la secuencia ancestral del ancestro respectivo se muestran en rojo, verde y negro respectivamente. **A.** Probabilidades más altas para la reconstrucción de secuencias de AncOP. **B.** Probabilidades para la reconstrucción de las secuencias de AncP.

Los dos clados que divergieron a partir de AncP en la subfamilia FoxP dieron lugar a proteínas con la capacidad de oligomerizar vía intercambio de segmentos, pero con dos tendencias claras: una en donde se estabilizó fuertemente el dímero con intercambio de segmentos (clado de FoxP3) y otra en donde la capacidad de dimerización se ha ido perdiendo a lo largo de la evolución (clado FoxP1, FoxP2, FoxP4) (Figura 22). Las características de la secuencia de AncP sugieren que puede dimerizar por intercambio de segmentos, y además que es una proteína que solo difiere en dos aminoácidos con FoxP1, por lo que puede tener propiedades de dimerización muy similares a FoxP1 o, considerando las mutaciones que el ancestro tiene respecto a la secuencia de FoxP1, tener propiedades de disociación más similares a FoxP3, lo que implicaría que dichas mutaciones tuvieran un efecto estabilizante en el dímero.

La cinética de disociación del dímero de AncP fue seguida a 37 °C al igual que las mutantes descritas previamente en esta tesis. AncP demostró tener una cinética de disociación relativamente similar a lo observado con FoxP1 (Figura 24), alcanzando el equilibrio dímero-monómero cerca de los 60 min. La  $K_D$  calculada para el ancestro fue de 2.4  $\mu$ M, cerca de 2 veces sobre la  $K_D$  de FoxP1, lo que en términos de energía libre de disociación ( $\Delta G_D$ ) se traduce en una diferencia de alrededor de 0,4 kcal/mol (Tabla 12). Sugiriendo que los residuos E69 y I52 no tienen un efecto importante en la dimerización por intercambio de segmentos. Sin embargo, este resultado indica que, al momento del origen del intercambio de segmentos en esta subfamilia, surgió con una capacidad considerablemente mayor a FoxP2 y FoxP4, pero menor a FoxP3. Esto sugiere que con el origen del intercambio de segmentos se dio a lugar a la formación de oligómeros en el contexto de una familia monomérica, pero que adquirieron con el curso de la evolución, capacidades de dimerización muy diferentes a partir del origen de AncP.



**Figura 24. Cinéticas de disociación del dímero de FoxP1 y el AncP.** El dímero aislado de FoxP1 y AncP y luego se siguió su cinética de disociación a 37 °C en experimentos de cromatografía de exclusión molecular. Los datos fueron ajustados a una función de decaimiento exponencial. A partir de la fracción al equilibrio obtenida de este ajuste y la constante cinética de disociación se obtiene la  $K_D$  reportada en la tabla 12.

**Tabla 12. Parámetros obtenidos a partir de las cinéticas de disociación de AncP y FoxP1.**

	$K_D$	$\Delta G_D$	$\Delta G_d^\ddagger$	$\Delta G_a^\ddagger$
	( $\mu M$ )	( $\mu M$ )	( $kcal \cdot mol^{-1}$ )	( $kcal \cdot mol^{-1}$ )
FoxP1	1,2 $\pm$ 0,20	8,4 $\pm$ 0,10	22,5 $\pm$ 0,10	14,1 $\pm$ 0,10
AncP	2,4 $\pm$ 0,05	8,0 $\pm$ 0,05	22,0 $\pm$ 0,10	14,0 $\pm$ 0,11

\*Las  $K_D$ , valores de energía libre y los respectivos errores provienen del ajuste de los datos de las cinéticas de disociación a una función de decaimiento exponencial, como se detalla en materiales y métodos.

## 4. DISCUSIÓN.

### 4.1. La barrera entálpica de asociación determina las diferencias en la dimerización de FoxP4 y FoxP1.

Los estudios de la cinética de disociación del dímero de FoxP4 (Figura 3) permitieron completar la caracterización de la subfamilia FoxP, la única subfamilia de los factores de transcripción Fox reportada con evidencia estructural de dimerización por intercambio de segmentos. En estos estudios se mostró que FoxP4 tiene constantes de disociación en el rango micromolar, aunque con valores considerablemente mayores a las constantes observadas para FoxP1 que a 33 °C tiene una  $K_D$  cerca de 135 veces más baja (Tabla 1). La  $K_D$  de FoxP4 se ubicó entre los valores de FoxP1 y FoxP2, siendo esta última la proteína con mayor tendencia a la disociación, con una  $K_D$  de cerca de 2 mM (Perumal et al., 2015). Esto implica que, agrupando por su capacidad de dimerizar, desde FoxP1 y finalizando con FoxP2, y considerando los resultados del análisis evolutivo de FoxP (Figura 22), la capacidad de formar el dímero en la rama que del árbol filogenético que agrupa a estas proteínas se ha ido perdiendo durante la evolución de esta subfamilia. Pese a ello, el intercambio de segmentos en esta subfamilia está altamente favorecido respecto a otros casos de proteínas que oligomerizan por este mecanismo, como el caso de p13suc1 (Rousseau et al., 2001) o cianovirina-N (Liu et al., 2012).

Los análisis de la cinética de disociación indican que en FoxP4 el intercambio de segmentos está termodinámicamente desfavorecido en  $3,0 \text{ kcal}\cdot\text{mol}^{-1}$  respecto a FoxP1 a 33 °C (Tabla 1). Además, el análisis de la dependencia de las constantes cinéticas con la temperatura (Figura 4 B) permitió determinar las barreras energéticas de asociación y disociación respecto al estado de transición en ambas proteínas. Esos análisis indican que las diferencias están mayormente en la reacción de asociación, en donde FoxP4 presenta una barrera energética ( $\Delta G_a^\ddagger$ )  $1,8 \text{ kcal}\cdot\text{mol}^{-1}$  superior a FoxP1. En esa línea, las mayores diferencias en las barreras entálpicas fueron observadas para el proceso de asociación ( $\Delta H_a^\ddagger$ ), donde FoxP4 presenta un  $\Delta H_a^\ddagger \sim 28 \text{ kcal}\cdot\text{mol}^{-1}$  mayor a lo observado para FoxP1 (Figura 4 C). Esto indica que, si bien ambas proteínas deben romper contactos para superar la barrera que separa al monómero con el estado de transición, la gran barrera entálpica de FoxP4 sugiere que esta proteína requiere una pérdida de contactos mayor para alcanzar el estado de transición, lo que explica su baja probabilidad de dimerizar cuando se le compara con FoxP1.

Pese a la gran barrera entálpica que FoxP4 tiene respecto a FoxP1, ambas barreras son considerablemente menores a las estudiadas para otras proteínas que dimerizan por intercambio de segmentos, como es el caso de la proteasa del virus SARS-CoV (Kang et al., 2012) que tiene  $\Delta H_a^\ddagger$  de 89,3 kcal·mol<sup>-1</sup> y otras que se encuentran en el mismo el mismo rango como estefina-A (Jerala & Zerovnik, 1999) y cianovirina-N (Liu et al., 2012). Esto sugiere que a pesar de la alta similitud de secuencia y estructura de ambas proteínas y en general de las proteínas de la subfamilia FoxP, los cambios que han dado origen a sus miembros han tenido consecuencias energéticas que generan propiedades de dimerización diferentes y que ellas están asociadas principalmente a las barreras entálpicas de asociación (Figura 4 C).

Por ende, las interacciones entre aminoácidos que se rompen y forman durante la asociación son claves para entender la diferencia en la barrera entálpica de transición entre ambas proteínas y el resto de la subfamilia FoxP. Al comparar sus secuencias, se observa que las diferencias más notables están distribuidas en varias regiones, excepto en la hélice H3 o las hebras S1, S2 y S3, que en mayor medida tienen una alta conservación de secuencia en toda la subfamilia (Figura 21). Algunas de las diferencias entre las secuencias de las proteínas FoxP fueron evaluadas mediante la sustitución de residuos en FoxP1 por aquellos observados en FoxP2 y FoxP4 (Figura 11 A y Figura 21). De esta forma se diseñaron dos mutantes de FoxP1, K23R y F80Y, y ambas mutantes tuvieron efectos modestos en la  $K_D$  (Tabla 5), sugiriendo que no son suficientes para dar cuenta de las diferencias observadas en la dimerización de estas proteínas. Otro cambio entre ambas proteínas, menos conservativo que los ensayados durante este trabajo de tesis, se observa muy próximo a la bisagra, en una región desestructurada previa al extremo amino terminal de la hélice H3. En FoxP1 se observa una alanina (A47) y en FoxP4 una treonina (T47) (Figura 21). La tendencia de estos residuos para encontrarse en la posición anterior en la secuencia del amino terminal de  $\alpha$ -hélices ha sido estudiada y se ha reportado que la presencia de treonina y alanina tienen tendencias diferentes según la posición que ocupen previo a  $\alpha$ -hélices (Kumar & Bansal, 1998), sugiriendo que ese cambio puede tener un efecto importante en la estructura de la hélice H3 y la región bisagra. Otro residuo que puede dar cuenta de las diferencias en la barrera entálpica de asociación de FoxP4 respecto a FoxP1 es R78, ubicado en la hélice H5 de FoxP4, el cual es único respecto a las otras secuencias de la subfamilia FoxP que presentan

residuos de carácter hidrofóbico en esta posición (Figura 21). Si bien la cadena lateral de R78 aparece de la hélice H5 está expuesto al solvente, interactúa con D76 de la misma cadena polipeptídica, localizándose este último en una región desestructurada antes del amino terminal de la hélice H5. Las simulaciones moleculares de los monómeros de FoxP4 (Figura 13) mostraron que el contacto formado entre R78 y D76 es estable durante gran parte de las simulaciones, sugiriendo que es un contacto nativo estable en FoxP4 y que en otros miembros de la familia no se aprecia.

Por otro lado, contrariamente a lo esperado, dos mutaciones que aumentaron 27 y 26 veces la constante de disociación en FoxP1 fueron Y40F y Q15W (Tabla 5), respectivamente. En el caso de la primera es aún más sorprendente, ya que la mutación en FoxP2 (Y40F), favoreció la formación del dímero cerca de 9 veces (Perumal et al., 2015). En el caso de Q15W, que a partir del análisis estructural del dímero de FoxP3 es descrita como clave para la formación de la interfaz secundaria formada entre las hélices H1 de ambas cadenas polipeptídicas (Bandukwala et al., 2011), tampoco favoreció el dímero. Ambos cambios forman parte de la interfaz secundaria que estabilizan la interfaz entre las hélices H1 del dímero de FoxP3 (Bandukwala et al., 2011). El hecho de que ambas mutantes favorezcan la disociación y por lo tanto la población monomérica, es entendible desde el punto de vista de que estas mutaciones están hechas en una proteína que precisamente, por las características de su secuencia, no posee una interfaz secundaria de asociación como la descrita en FoxP3. En efecto, una mutante triple que incluye la mutación reversa a la hecha en FoxP1 (Q15W) en FoxP3 (W15Q/M37T/A39P), en donde los dos primeros reemplazos corresponden a mutaciones de la interfaz secundaria y la última el cambio en la región bisagra que vuelve a FoxP1 (Chu et al., 2011) (Medina et al., 2016) y FoxP2 (Stroud et al., 2006) completamente monoméricas, no permitió convertir a FoxP3 en una proteína completamente monomérica (Bandukwala et al., 2011). Esto sugiere que los cambios necesarios para optimizar y favorecer el dímero por intercambio de segmentos son más complejos, y, probablemente, pueden involucrar varias mutaciones simultáneas que ocurrieron durante la evolución de esta familia y que actúan de manera sinérgica en la estabilización de la oligomerización.

Este análisis sugiere que las diferencias en barreras energéticas de asociación observadas en FoxP4 y FoxP1, y probablemente en toda la subfamilia, no son explicables

por cambios únicos en la secuencia, sino que más bien debieron ocurrir de forma concertada entre dos o más grupos de residuos, de esta forma modelando las propiedades de dimerización de todos los miembros de la subfamilia.

#### **4.2. Un intermediario monomérico favorece la formación del intercambio por segmentos en la familia FoxP.**

El intercambio de segmentos ha sido definido como un mecanismo de oligomerización que requiere eventos de desplegamiento para asociar dos o más cadenas polipeptídicas. El estudio comparativo de las vías de plegamiento y estabilidad termodinámica de los miembros de la subfamilia FoxP mostró diferencias importantes entre sus miembros (Figura 7 y Tabla 3). En FoxP2 y FoxP4, en las que el estado monomérico de ambas proteínas estaba altamente favorecido, el mecanismo que mejor explicó el desplegamiento de las proteínas fue uno de dos estados ( $N_2 \rightleftharpoons 2U$ ). Por otra parte, aquellas proteínas en donde a las condiciones utilizadas el dímero fue la especie favorecida (FoxP1 y FoxP3), el mecanismo que mejor describió el desplegamiento fue uno de tres estados con un intermediario monomérico ( $N_2 \rightleftharpoons 2I \rightleftharpoons 2U$ ).

Los resultados del desplegamiento al equilibrio de FoxP4 y FoxP2 son interesantes desde la perspectiva en que ambas proteínas son capaces de oligomerizar por medio del intercambio de segmentos, pero las condiciones ensayadas favorecen la presencia del monómero. Los resultados indican que ambas proteínas tienen un mecanismo de plegamiento de dos estados, pero con una estabilidad termodinámica que difiere en 2,3 kcal·mol<sup>-1</sup>, siendo el monómero de FoxP2 el más estable termodinámicamente entre ambas, lo cual es consecuente con su baja tendencia a encontrarse como dímero respecto a los otros miembros de la subfamilia (Perumal et al., 2015).

En base a estos resultados, el mecanismo de plegamiento de dos estados parece ser un mecanismo conservado en los monómeros de la familia Fox. A esto se suma que la introducción de una prolina, observada en todas las proteínas Fox monoméricas, en reemplazo de alanina en la región bisagra, vuelve a FoxP1 completamente monomérica y su mecanismo de plegamiento fue descrito como uno de dos estados al seguir el desplegamiento por dicroísmo circular (Medina et al., 2016), tal como lo descrito en esta tesis para FoxP2 y FoxP4 silvestre en condiciones en donde el monómero de ambas proteínas está favorecido

(Figura 5 A). Esto sugiere fuertemente que el mecanismo de tres estados es una novedad evolutiva dentro de la familia y que surgió con la aparición de la primera proteína capaz de dimerizar en la familia, AncP (Figura 24)

Respecto al mecanismo de desplegamiento en tres estados  $N_2 \rightleftharpoons 2I \rightleftharpoons 2U$ , descrito para FoxP1 y FoxP3, la etapa de disociación fue relacionada con la primera transición, ya que ha demostrado ser dependiente de la concentración de proteína tanto para FoxP1 (Medina et al., 2016) como para FoxP3 (Figura 7 C). En efecto, la diferencia de energía libre calculada para la disociación en los experimentos de cinéticos a 37 °C de Foxp1( $\Delta G_D$ ) (Tabla 1) y la obtenida en los estudios de desplegamiento al equilibrio ( $\Delta G_U$ ) para esa proteína (Tabla 3) difieren solo en  $\sim 0,1 \text{ kcal}\cdot\text{mol}^{-1}$ . Además, los experimentos permitieron establecer que la disociación involucra cerca de 20% de pérdida del contenido de estructura secundaria, manteniendo a los monómeros parcialmente desplegados luego de disociarse.

De forma opuesta a lo observado en FoxP1 y FoxP3, en otros modelos como p13suc1, la disociación y el desplegamiento total de la proteína involucran cambios en la energía libre de magnitudes similares, lo que indica que p13suc1 requiere transitar por un estado completamente desplegado para oligomerizar por intercambio de segmentos (Rousseau et al., 2001). En ese aspecto, el hecho de que las proteínas FoxP requieran perder solo parte de su estructura nativa para formar un estado intermediario apto para el intercambio de segmentos da luces de por qué este mecanismo de asociación en esta subfamilia se da en condiciones en ausencia de agentes caotrópicos que perturben el estado nativo, resaltando de esta forma la relevancia biológica de este mecanismo en las proteínas FoxP.

El análisis de la energía libre involucrada en la disociación indica que la diferencia entre FoxP1 y FoxP3 para ese paso ( $N_2 \rightleftharpoons 2I$ ) es de  $\sim 2,5 \text{ kcal}\cdot\text{mol}^{-1}$ , indicando que la disociación está desfavorecida en FoxP3 respecto a FoxP1. Aunque para FoxP3 se carece de información de la estructura del intermediario, se ha indagado en los arreglos estructurales que involucra formar un intermediario monomérico en FoxP1 mediante intercambio de hidrógeno por deuterio seguido por espectroscopia de masas (HDX-MS) y cromatografía de exclusión molecular en condiciones que favorecen la formación de esta especie (2 M de GndCl) (Medina et al., 2016). Los análisis de cromatografía de exclusión molecular en esas condiciones indicaron que el intermediario corresponde a una conformación de un monómero

expandido y los experimentos de HDX-MS que la interacción entre las hebras S1 y S3 se pierden, quedando ambas regiones más expuestas al solvente que lo observado en el monómero nativo (Medina et al., 2016). Esto es interesante, porque la interacción entre ambas hebras es parte de la interfaz que debe romper sus contactos en el monómero y que debe reestablecerse en el dímero de forma intermolecular, sugiriendo que las diferencias entre ambas proteínas en esta región pueden ser claves para modelar las propiedades de oligomerización en ambas proteínas. En ese sentido, la comparación de las secuencias de esas regiones muestra que en la hebra S1 hay una gran conservación de secuencia en las proteínas FoxP, excepto en un residuo: mientras que FoxP1 y el resto de la subfamilia tienen un residuo de carácter hidrofóbico como leucina, FoxP3 tiene en la posición equivalente un residuo cargado positivamente de arginina (Figura 21). En las estructuras disponibles para las proteínas de la subfamilia, la cadena lateral de ese residuo se orienta entre las hélices H2 y H1, sugiriendo que las propiedades fisicoquímicas de la cadena lateral puedan influir en el empaque de ambas hélices y favorecer las diferencias en la tendencia a formar este intermediario monomérico. Adicionalmente, es importante destacar que gran parte de la escasa variabilidad de secuencia entre las proteínas FoxP se encuentra precisamente alrededor de esta región, entre la mitad del carboxilo terminal de H1, S1 y H2/H4 (Figura 21), por lo que es esperable que otros residuos en esta o regiones contiguas también tengan un gran impacto en la tendencia a dimerizar en la subfamilia FoxP y, consecuentemente, en la formación del intermediario monomérico.

Estos antecedentes indican que la capacidad de dimerizar por intercambio de segmentos está estrechamente relacionada con la presencia de un intermediario monomérico en las proteínas FoxP. La desestabilización del monómero durante la evolución tiene directa relación con la aparición de la mutación de prolina por alanina en estas proteínas (Figura 21 y Figura 22), pero los cambios de secuencia que modelaron las diferencias entre las proteínas FoxP parecen afectar directamente a la estabilidad del intermediario y concomitantemente a la capacidad de dimerizar por intercambio de segmentos. Esta capacidad de dimerizar por medio de la formación de un intermediario monomérico solo parcialmente desestructurado y no por vía del desplegamiento total de la proteína explica por qué el intercambio de segmentos está fuertemente favorecido en comparación a los ejemplos de intercambio de segmentos ya mencionados.

### **4.3 Los miembros de la subfamilia FoxP poseen una dinámica conformacional diferente.**

La capacidad de dimerizar por intercambio de segmentos de la subfamilia FoxP está estrechamente relacionada con la capacidad de formar un intermediario monomérico, el que respecto al monómero nativo pierde parte de su contenido de estructura nativa (Figura 7 B). En ese aspecto, la obtención de la estructura de FoxP4 completó la historia estructural de los monómeros de esta subfamilia, sumándose a las estructuras disponibles para el monómero de FoxP2 (Stroud et al., 2006) y el de la mutante de FoxP1 (A39P/C67Y) (Chu et al., 2011). La estructura del monómero de FoxP4 (Figura 6) careció de densidad electrónica suficiente para resolver la hélice H5, lo que sugiere una alta heterogeneidad conformacional de esa región. De acuerdo con ese antecedente, esa región ha sido sugerida como altamente dinámica en FoxP1 a partir del análisis estructural de las conformaciones obtenidas por RMN (Chu et al., 2011), en acuerdo con las simulaciones moleculares de FoxP1 y FoxP4 (Figura 12).

Los estudios publicados respecto a FoxP1 demuestran que el intermediario monomérico es expandido y que la apertura de ese monómero es producto de la pérdida de la interfaz entre las hebras S1-S3 (Medina et al., 2016). Para que este evento ocurra, es prerequisite que los contactos que estabilizan la interacción entre las hélices H1 y H5 se pierdan. Esto sugiere que la formación del dímero con intercambio de segmentos requiere de una alta heterogeneidad conformacional de la estructura de los miembros de la familia FoxP cuya oligomerización se vea favorecida termodinámicamente. En ese sentido, se realizaron simulaciones moleculares de 180 ns para evaluar la dinámica conformacional del estado nativo de las estructuras disponibles para los miembros de la subfamilia. En el caso de los monómeros, en el corto tiempo simulado solo fue posible observar, en parte de las simulaciones, la pérdida de interfaz H1-H5 en FoxP1 y FoxP4 (Figura 12 y Tabla 6), mientras que en FoxP2 esa interacción permaneció estable, como indicó el análisis de la interacción por puente de hidrógeno (Figura 13) (Figura 12 y Tabla 6). Ese antecedente junto a la evidencia experimental que demuestra que en el intermediario monomérico la interfaz formada entre las hélices S1 y S3 se pierde (Medina et al., 2016) (Medina et al., 2020), permiten sugerir que en esta región ocurre uno de los primeros eventos que conllevan a la desestabilización del monómero y la oligomerización los miembros de la subfamilia FoxP.

Los eventos que gatillen un cambio conformacional suficientemente grande como para desestabilizar el monómero nativo compacto y completamente plegado, deben originarse a partir de eventos de desestabilización local de regiones como la bisagra y otras que permitan la apertura del monómero. En análisis de los movimientos correlaciones de las proteínas FoxP, destacan algunos movimientos correlacionados de cadenas laterales que sugieren movimientos colectivos en la estructura de los monómeros. Es así como el residuo R14 de la hélice H1, que aporta con su cadena lateral para estabilizar H5 en la estructura de FoxP2, está muy correlacionado con residuos del lazo entre S1 y la región del carboxilo terminal de H3 (Figura 14 y Tabla 8). Esta vía de comunicación es interesante porque aparece muy favorecida en FoxP1 y FoxP4, pero no en FoxP2. En acuerdo con esos antecedentes, en el extremo carboxilo de la hélice H3, la protonación de H59 ha demostrado desfavorecer el intercambio de segmentos por medio del aumento de las barreras cinéticas de asociación (Medina et al., 2019). A nivel atómico ese efecto fue explicado en simulaciones de plegamiento que mostraron que la desprotonación de H59 favorece contactos en el estado de transición de plegamiento entre los residuos de H3-S3 y H1 (Medina et al., 2019). Además, la mutación C61Y localizada entre H3 y S2 en FoxP1 aumentó la constante de disociación 12 veces en esa proteína (Chu et al., 2011). Por consiguiente, estas regiones pueden ser importantes para estabilizar el intermediario monomérico y por lo tanto para la dimerización.

Las simulaciones moleculares de los dímeros de las proteínas con tendencias a dimerizar más opuestas en la subfamilia, correspondientes a FoxP2 y FoxP3, sugieren que hay importantes diferencias en la dinámica conformacional de ambas proteínas. FoxP3 es un dímero compacto como lo sugieren los análisis de radio de giro (Fig 19.D) y las distancias entre los residuos W15 en ambas hélices H1 del dímero (Fig 15.C), y, por otro lado, FoxP2 explora conformaciones más expandidas como lo indica sus radios de giro (Fig 19.D) y la distancia entre los residuos Q15 equivalentes a W15 en FoxP3 (Figura 19 C). Estas diferencias conformacionales serían una consecuencia del desarrollo de una interfaz secundaria, adicional a la formada por la bisagra, una característica propia de algunas proteínas que oligomerizan por intercambio de segmentos y que, luego del cambio en su región bisagra, durante la evolución han desarrollado esta interfaz secundaria que permite estabilizar al dímero con intercambio de segmentos. Las simulaciones moleculares convencionales en solvente explícito, usando el estado nativo de estas proteínas como sistema

inicial, permiten evaluar la dinámica conformacional de este estado, pero limitada a cambios conformacionales pequeños. En ese sentido, las simulaciones de plegamiento de FoxP1, que fueron incorporadas en un estudio de FRET de molécula única recientemente publicado (Medina et al., 2020), sirvieron para explorar un espacio conformacional más amplio que las simulaciones convencionales. Las distancias estimadas a partir del enfoque de FRET de molécula individual usando fluoróforos posicionados en diferentes localizaciones de la proteína fueron bien representadas por el ensamble nativo del dímero (N) obtenido en las simulaciones de plegamiento (Figura 19). Esto provee de evidencia en solución de la existencia del intercambio de segmentos en la subfamilia FoxP, en adición a la información derivada de las estructuras cristalográficas disponibles para la familia. Por otra parte, dos intermediarios fueron identificados en las simulaciones, lo que sugiere que el dímero de FoxP1 previo al desplegamiento se conecta al menos con dos ensamblajes conformacionales (I1 e I2) que se encuentran parcialmente desplegados pero que retienen contactos que lo hacen preservar su estado dimérico con intercambio de segmentos.

La flexibilidad conformacional y la desestructuración mostrada por el dímero de FoxP1 puede ser una ventaja adaptativa respecto a proteínas con un dímero más rígido y estable, lo que le permite interactuar con otras cadenas polipeptídicas en escalas de tiempo pequeñas. Esto clasifica a FoxP1 como una proteína intrínsecamente desordenada, al menos en su conformación dimérica. El alto grado de desestructuración de este tipo de proteínas le permite interactuar con otras macromoléculas como ADN y/u otra cadena polipeptídica por medio de la estabilización de estados correctamente plegados (estructurados). De esta forma, es común que las proteínas intrínsecamente desordenadas acoplen la unión y el plegamiento en una nueva conformación al asociarse con otra macromolécula en lo que ha sido propuesto como el mecanismo de *fly-casting* (Y. Huang & Liu, 2009). En el caso de factores de transcripción, se ha propuesto que este mecanismo permite a la proteína unirse al ADN en un estado parcialmente plegado de manera inespecífica y luego plegarse por completo cuando se une a su sitio de unión en el ADN (Levy et al., 2007).

#### **4.4. El intercambio de segmentos es modulado por residuos fuera de la región bisagra.**

Los estudios sobre el efecto del pH en la dimerización por intercambio de segmentos muestran que el pH tiene un efecto en el equilibrio de dimerización y que ese efecto sería

explicado por la desprotonación de H59 (Medina et al., 2019). Estudios complementarios de anisotropía usando sondas fluorescentes ligadas a la hélice H3 sugieren que ese efecto se lograría porque la desprotonación provoca la desestabilización local de esa hélice (Medina et al., 2019). A nivel atómico, y como parte de los resultados de esta tesis que forman parte del artículo citado anteriormente, se estudió el efecto de la protonación en la estructura de FoxP1 con simulaciones moleculares. En estas simulaciones se aprecia que la desprotonación de la histidina favorecería la formación de contactos de la cadena lateral de ese residuo con la cadena principal de N55 (Figura 10). Además, simulaciones moleculares de plegamiento del monómero mostraron que la desprotonación de H59 favorecería la estabilización de interacciones nativas en el estado de transición, como las observadas entre F62 (hebra S2) y R14(hélice H1) y la interacción entre H59 y N55.

El efecto del estado de protonación para un residuo como histidina no es sorprendente ya que este residuo ha sido relacionado tanto al plegamiento de proteínas como a su función (Liao et al., 2013)(Connelly & McIntosh, 1998). El efecto estabilizante de las cadenas laterales en posiciones en la estructura más allá en la secuencia del extremo amino y carboxilo terminal de alfa hélices ha sido descrito como relevantes para la estabilidad de las estructuras de  $\alpha$ -hélices (Kumar & Bansal, 1998). Esto, porque forman interacciones de puente de hidrógeno con las amidas y carbonilos de la cadena principal que quedan disponibles por encontrarse en los extremos de las hélices (Kumar & Bansal, 1998). A nivel estructural, esto explica el efecto estabilizante que tendría la desprotonación de la H59 en la hélice H3 de FoxP1 y que también ocurriría en otras FoxP, dada la alta similitud estructural y de secuencia en esta región para sus distintos miembros.

En el contexto de la familia Fox, la posición de H59 en una parte importante de sus miembros tienen una asparagina (N55) (Figura 21), que de igual forma puede aportar con su cadena lateral para establecer puentes de hidrógeno con la cadena principal de H3. Sin embargo, histidina, observada solo en los homólogos cercanos a FoxP (FoxO y FoxM) (Figura 21) tiene la posibilidad de actuar como un sensor molecular al variar su estado de protonación según el contexto estructural en el que esta cadena lateral se encuentre, producto de que los residuos de histidina poseen un pKa que actúa en rangos de pH de interés fisiológico como ha sido demostrado para la función de proteínas (Liao et al., 2013) (Zhu

et al., 2017) (Park & Raines, 2000). En ese aspecto, al observar el árbol filogenético (Figura 22) se aprecia que la aparición de esta histidina es previa al origen de la familia FoxP; probablemente apareció por primera vez en la familia entre el último ancestro en común con la familia FoxO, AncOP, y el resto de las proteínas monoméricas, ya que la secuencia calculada de ese ancestro tiene una histidina en esa posición y no asparagina (Figura 21).

También, alejado de la región bisagra, el residuo F7 se localiza en la hélice H1 y fue identificado como un residuo PCAIN. Este residuo muestra un cambio conformacional importante al comparar su conformación en el monómero y en el dímero. En el monómero se encuentra expuesto al solvente y en el dímero forma parte de una red de interacciones dominada por residuos aromáticos que estabilizan el dímero (Figura 8). Para evaluar la importancia de ese residuo, se reemplazó por una leucina observada en la posición equivalente en la subfamilia FoxO, la más cercana filogenéticamente con la subfamilia FoxP. La caracterización de la cinética de la disociación del dímero y el posterior análisis indicó que la mutante genera un aumento de la constante de disociación (24 veces) y la barrera cinética de asociación (Figura 9 y Tabla 4). Este resultado es consistente con el rol que se ha asociado a residuos aromáticos de fenilalanina como importante para la formación de interfaces proteína-proteína (Ma & Nussinov, 2007).

El efecto de H59 que puede modular el intercambio de segmentos según su estado de protonación y F7 que participa de la red de interacciones que favorecen la formación del dímero sugiere que ambos residuos, que no están conservados en toda la familia Fox, pero que surgieron previo a la mutación de la región bisagra P39A, y que se encuentran ubicados en zonas alejadas de la región bisagra indica que el intercambio de segmentos puede ser modulado por aminoácidos fuera de esa región, tal como se demostró para p13suc1 (Schymkowitz et al., 2001).

#### **4.5. Las dos ramas filogenéticas de la subfamilia FoxP formaron dímeros con propiedades de oligomerización diferentes.**

El árbol filogenético de la familia Fox (Figura 22) sugiere cómo han evolucionado estos factores de transcripción, indicando que la subfamilia FoxP emergió a partir de un ancestro en común con la subfamilia FoxO. Estas últimas son proteínas que han sido descritas como monoméricas en todas las estructuras conocidas a la fecha (Stroud et al., 2006) (Chu et al., 2011) (Bandukwala et al., 2011), al igual que muchos miembros del resto de la familia. La excepción la hace la familia FoxP, donde el último ancestro en común para sus miembros, AncP, fue descrito con una  $K_D$  similar a la calculada para FoxP1 (Fig 24 y Tabla 12). En efecto, el ancestro AncP sería la primera proteína en la historia evolutiva de esta familia que formaría oligómeros por intercambios de segmentos, ya que el último ancestro en común para la familia O y P, AncOP, posee prolina en la región bisagra, la que es característica de las proteínas monoméricas en esta familia y sugeriría por tanto su impedimento de dimerizar (Stroud et al., 2006).

Todos los estudios disponibles indican que la mutación de alanina por prolina en la región bisagra es la gran responsable del origen del intercambio de segmentos (Stroud et al., 2006)(Chu et al., 2011)(Medina et al., 2016) lo que se ve observado al menos tanto para FoxP1 como para FoxP2. Sin embargo, desde la emergencia de AncOP a AncP, y adicional a la mutación en la región bisagra, ocurrieron otros cambios de secuencias como se aprecia en el alineamiento de secuencias de la familia (Figura 21). Respecto a las proteínas FoxP y su ancestro AncP, el ancestro AncOP tiene algunas inserciones en lazo 1 (W1) y la hélice H5. La inserción en el lazo W1 ha sido estudiada en nuestro laboratorio (datos no publicados), determinándose que la mutante tiene una constante de disociación similar a la observada en FoxP1. Por otra parte, la flexibilidad conformacional de la hélice H5 observada en las simulaciones moleculares del dímero de FoxP3 (Figura 15 B) o los monómeros de FoxP1 y FoxP4 (Figura 12 B y 12 C), indican que esa región se altamente dinámica e incluso pierde su estructura, en el caso del dímero de FoxP3. Por lo mismo, es probable que una inserción en esa región, tal como la inserción PG en la aleta W1 observada en el ancestro AncOP, no tenga un efecto en la magnitud del efecto de la mutación en la región bisagra. Por ello, es esperable que las mutaciones que ocurrieron junto a A39P en la trayectoria evolutiva entre

AncOP y AncP sean accesorias o secundarias en el contexto de esa mutación para modular el intercambio de segmentos, como los casos descritos en la sección anterior de F7 y H59.

A partir del ancestro AncP emergen dos clados: uno de ellos al que pertenece FoxP3 y el otro a FoxP1, FoxP2 y FoxP4 (Figura 22). Ambos clados representan tendencias de dimerización diferentes. Por un lado, el clado que muestra exclusivamente a FoxP3 contiene a la proteína caracterizada como un dímero obligado y la más estable termodinámicamente de la subfamilia FoxP (Tabla 3). Por otro, el clado que reúne a los demás miembros, muestra que, desde FoxP1-el más próximo a AncP- hasta FoxP2 -el más alejado a AncP- la  $K_D$  varía desde el orden micromolar (Tabla 1), al orden milimolar (Perumal et al., 2015). Esta tendencia opuesta, abre la interrogante de por qué y cómo surgen tendencias tan disímiles a partir de proteínas que divergieron a partir de un ancestro en común y entre ellas tienen un alto grado de identidad de secuencia (84% identidad de secuencia).

Como se comentó anteriormente, las mutaciones que originaron el intercambio de segmentos a lo largo de la evolución de la familia Fox permitieron desestabilizar solo parcialmente el monómero, posiblemente comportándose como elementos intrínsecamente desordenados, lo que explica el mecanismo de plegamiento de tres estados observado para FoxP1 y FoxP3. Esto marca una gran diferencia con modelos canónicos de intercambio de segmentos como p13suc1 (Rousseau et al., 2001) o la cianovirina (Liu et al., 2012), que tienen un mecanismo de dos estados, en donde la única forma de plegarse como dímero con intercambio de segmentos es vía desplegamiento completo de los monómeros y la acción cooperativa del plegado de las dos cadenas polipeptídicas. Sin embargo, para las proteínas FoxP, esto también tiene como consecuencia que las mutaciones que se originaron a partir de AncP que podrían eventualmente estabilizar el intermediario monomérico o estabilizar el monómero nativo y compacto, como se observa para la rama de FoxP3 y de FoxP1-FoxP2-FoxP4, respectivamente.

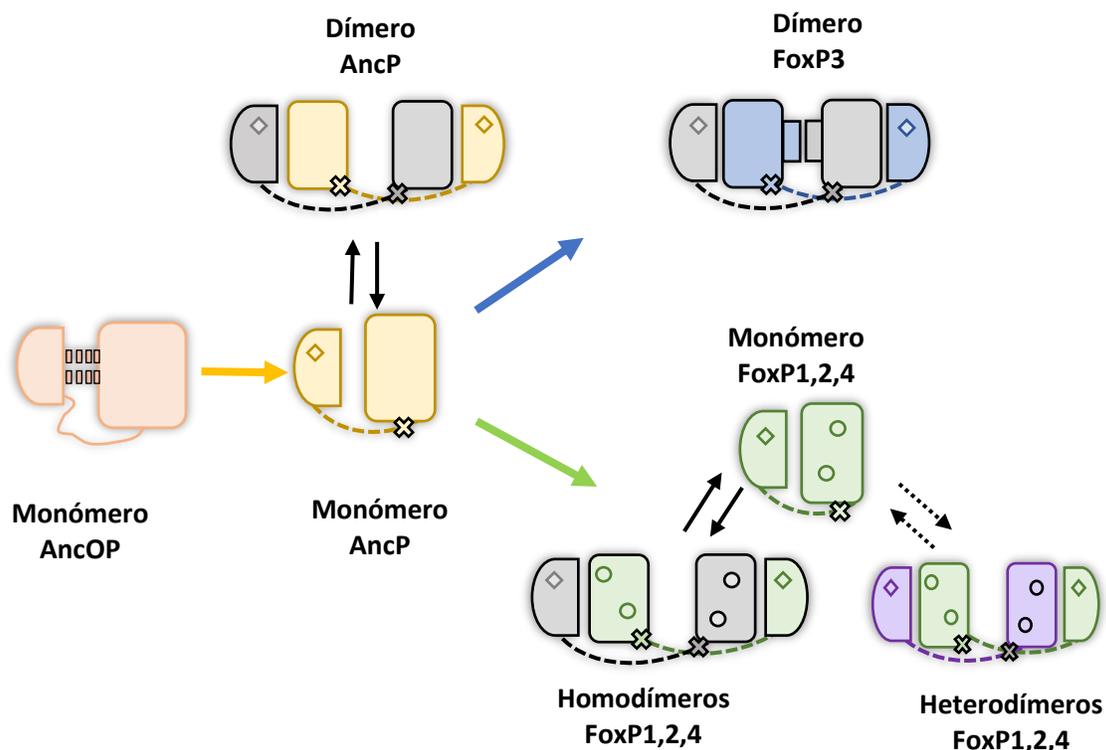
Para que las mutaciones que dan origen a las proteínas FoxP se fijen en la trayectoria evolutiva de la subfamilia deben tener también consecuencias funcionales. Es así como ha sido demostrado que el intercambio de segmentos en FoxP3 permitiría fomentar los contactos inter-cromosomales alrededor de un loci conocido de esta proteína en experimentos de secuenciación 4C-Seq (Bandukwala et al., 2011). Si bien esta consecuencia funcional

también puede ser considerada como una ventaja evolutiva para la otra rama evolutiva de la subfamilia que tiene una capacidad de dimerización disminuida (rama FoxP1-FoxP2-FoxP4), también es necesario considerar que la capacidad de oligomerizar por este mecanismo es variable entre estos miembros (Figura 22).

En ese caso se abre la pregunta de por qué, funcionalmente, estas proteínas podrían volver a estabilizar un estado monomérico. Una explicación probable es que ha sido demostrado que FoxP1, FoxP2 y FoxP4 se co-expresan en células del cerebro (Lu et al., 2002)(Teufel et al., 2003) mientras que FoxP3 es preferentemente expresado en células T del sistema inmune (Huehn et al., 2009). Consecuentemente con su co-expresión, recientemente se ha demostrado que FoxP1, FoxP2 y FoxP4 pueden heterodimerizar *in vivo* en células del área X del cerebro (Mendoza et al., 2017). Trabajos preliminares de nuestro grupo de laboratorio confirman esta observación luego de que con experimentos *in vitro* se demostrara que FoxP1 y FoxP2 heterodimeriza. Esto sugiere que las mutaciones ocurridas en la rama FoxP1-FoxP2-FoxP4 desfavorecerían la homodimerización y favorecerían la heterodimerización, tomando ventaja de la alta similitud de secuencia y estructural de estas proteínas.

## 5. CONCLUSIÓN

Los antecedentes expuestos en este trabajo sugieren que el origen del intercambio de segmentos en esta familia, al igual que en todas las proteínas que oligomerizan por este mecanismo, surgió por una mutación primaria en la región bisagra (A39P). En conjunto a esa mutación durante la evolución de la familia se habrían acumulado otros cambios de secuencias, considerados secundarios respecto a la mutación A39P y localizados en regiones alejadas de la bisagra (F7L y H59, entre otras), que afectan la estabilidad de la dimerización. La mutación primaria A39P permitió esto por medio de la desestabilización del monómero de la familia Fox, generando una especie monomérica parcialmente plegada que estuviera habilitada para asociarse con otra cadena polipeptídica. Estos y otros cambios originaron a la primera proteína con capacidad de oligomerizar, AncP, y a partir de este ancestro emergieron dos ramas evolutivas con tendencias de dimerización diferentes: la rama evolutiva de FoxP3, que por medio de la formación de una interfaz secundaria observada en el dímero forma una especie altamente estable conformacionalmente y termodinámicamente; y la rama filogenética que agrupa a FoxP1, FoxP2 y FoxP4, las cuales son más inestables que FoxP3, lo que ha traído como consecuencia que el monómero vuelva a ser favorecido en estas proteínas. Sin embargo, la pérdida de la capacidad de homodimerizar podría estar directamente implicada con la capacidad de heterodimerizar de los miembros de esta rama evolutiva de la familia (Figura 25).



**Figura 25. Origen y evolución del intercambio de segmentos en la familia Fox.** Los colores en los esquemas representan una cadena polipeptídica y las mutaciones son representadas con un círculo, rombo y cruz. A partir del ancestro AncOP, surgió la mutación P39A en la región bisagra que ha sido descrita como fundamental para el origen de este mecanismo (cruz), junto a ella aparecieron otras mutaciones como F7 y H59 (rombo en AncP) que contribuyen a la estabilización de un dímero con intercambio de segmentos. A partir de AncP emergieron dos ramas filogenéticas con capacidad de dimerizar por este mecanismo, pero con una probabilidad de formación del dímero muy diferente: la rama evolutiva de FoxP3 y la rama de FoxP1, FoxP2 y FoxP4. En la primera, los residuos en la hélice H1 como W15, F40, M37 entre otras (rectángulo pequeño entre ambas subunidades) desarrollaron una superficie de interacción secundaria, adicional a la bisagra que estabiliza fuertemente el dímero con intercambio de segmentos. En la otra rama evolutiva, surgieron aminoácidos (círculos) que hacia los miembros más actuales como FoxP2 y FoxP4 han desfavorecido la formación del homodímero con intercambio de segmentos. Esa desestabilización podría ser relacionada con la capacidad de heterodimerizar de FoxP1, FoxP2 y FoxP4 (flechas punteadas entre el monómero y heterodímeros).

## 6. REFERENCIAS

- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2). <https://doi.org/10.1107/S09074444909052925>
- Alva, V., Ammelburg, M., Söding, J., & Lupas, A. N. (2007). On the origin of the histone fold. *BMC structural biology*, 7(1), 17. <https://doi.org/10.1186/1472-6807-7-17>
- Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology*, 55(4). <https://doi.org/10.1080/10635150600755453>
- Bandukwala, H. S., Wu, Y., Feuerer, M., Chen, Y., Barboza, B., Ghosh, S., Stroud, J. C., Benoist, C., Mathis, D., Rao, A., & Chen, L. (2011). Structure of a Domain-Swapped FOXP3 Dimer on DNA and Its Function in Regulatory T Cells. *Immunity*, 34(4), 479–491. <https://doi.org/10.1016/j.immuni.2011.02.017>
- Bateman, A. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, 47(D1). <https://doi.org/10.1093/nar/gky1049>
- Bennet, M. J., Schlunegger, M. P., & Eisenberg, D. (1995). 3D Domain swapping: A mechanism for oligomeric assembly. *Protein Science*, 4, 2455–2468. <https://doi.org/10.1002/pro.5560041202>
- Bergdoll, M., Remy, M. H., Cagnon, C., Masson, J. M., & Dumas, P. (1997). Proline-dependent oligomerization with arm exchange. *Structure*. [https://doi.org/10.1016/S0969-2126\(97\)00196-2](https://doi.org/10.1016/S0969-2126(97)00196-2)
- Blane, A., & Fanucchi, S. (2015). Effect of pH on the Structure and DNA Binding of the FOXP2 Forkhead Domain. *Biochemistry*, 54(25), 4001–4007. <https://doi.org/10.1021/acs.biochem.5b00155>

- Bordo, D., & Argos, P. (1994). The role of side-chain hydrogen bonds in the formation and stabilization of secondary structure in soluble proteins. *Journal of Molecular Biology*. <https://doi.org/10.1006/jmbi.1994.1676>
- Borgia, A., Kemplen, K. R., Borgia, M. B., Soranno, A., Shammas, S., Wunderlich, B., Nettels, D., Best, R. B., Clarke, J., & Schuler, B. (2015). Transient misfolding dominates multidomain protein folding. *Nature Communications*. <https://doi.org/10.1038/ncomms9861>
- Brent, M. M., Anand, R., & Marmorstein, R. (2008). Structural Basis for DNA Recognition by FoxO1 and Its Regulation by Posttranslational Modification. *Structure*, 16(9). <https://doi.org/10.1016/j.str.2008.06.013>
- Brummer, T., & McInnes, C. (2020). RAF kinase dimerization: implications for drug discovery and clinical outcomes. En *Oncogene*. <https://doi.org/10.1038/s41388-020-1263-y>
- Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). The Amber biomolecular simulation programs. En *Journal of Computational Chemistry* (Vol. 26, Número 16). <https://doi.org/10.1002/jcc.20290>
- Chen, X., Wei, H., Li, J., Liang, X., Dai, S., Jiang, L., Guo, M., Qu, L., Chen, Z., Chen, L., & Chen, Y. (2019). Structural basis for DNA recognition by FOXC2. *Nucleic Acids Research*, 47(7). <https://doi.org/10.1093/nar/gkz077>
- Chen, Y., Chen, C., Zhang, Z., Liu, C.-C., Johnson, M. E., Espinoza, C. A., Edsall, L. E., Ren, B., Zhou, X. J., Grant, S. F. A., Wells, A. D., & Chen, L. (2015). DNA binding by FOXP3 domain-swapped dimer suggests mechanisms of long-range chromosomal interactions. *Nucleic acids research*, 43(2), 1268–1282. <https://doi.org/10.1093/nar/gku1373>
- Chen, Y. W., Stott, K., & Perutz, M. F. (1999). Crystal structure of a dimeric chymotrypsin inhibitor 2 mutant containing an inserted glutamine repeat. *Proceedings of the National Academy of Sciences of the United States of America*, 96(4), 1257–1261. <https://doi.org/10.1073/pnas.96.4.1257>
- Chu, Y.-P., Chang, C.-H., Shiu, J.-H., Chang, Y.-T., Chen, C.-Y., & Chuang, W.-J. (2011). Solution structure and backbone dynamics of the DNA-binding domain of FOXP1: Insight

into its domain swapping and DNA binding. *Protein Science*, 20(5), 908–924. <https://doi.org/10.1002/pro.626>

Chu, Y. P., Chang, C. H., Shiu, J. H., Chang, Y. T., Chen, C. Y., & Chuang, W. J. (2011). Solution structure and backbone dynamics of the DNA-binding domain of FOXP1: Insight into its domain swapping and DNA binding. *Protein Science*, 20(5), 908–924. <https://doi.org/10.1002/pro.626>

Clark, K L, Halay, E. D., Lai, E., & Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature*, 364(6436), 412–420. <https://doi.org/10.1038/364412a0>

Clark, Kirk L., Halay, E. D., Lai, E., & Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature*. <https://doi.org/10.1038/364412a0>

Connelly, G. P., & McIntosh, L. P. (1998). Characterization of a buried neutral histidine in *Bacillus circulans* xylanase: Internal dynamics and interaction with a bound water molecule. *Biochemistry*. <https://doi.org/10.1021/bi972085v>

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btr088>

DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. *{CCP4} Newsletter On Protein Crystallography*, 40.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4). <https://doi.org/10.1107/S0907444910007493>

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography*, 69(7). <https://doi.org/10.1107/S0907444913000061>

Eyring, H. (1935). The Activated Complex and the Absolute Rate of Chemical Reactions. *Chemical Reviews*, 17(1), 65–77. <https://doi.org/10.1021/cr60056a006>

Gammons, M. V., Renko, M., Johnson, C. M., Rutherford, T. J., & Bienz, M. (2016). Wnt Signalosome Assembly by DEP Domain Swapping of Dishevelled. *Molecular Cell*. <https://doi.org/10.1016/j.molcel.2016.08.026>

Goodsell, D. S., & Olson, A. J. (2000). Structural symmetry and protein function. En *Annual Review of Biophysics and Biomolecular Structure*. <https://doi.org/10.1146/annurev.biophys.29.1.105>

Goyal, B., & Goyal, D. (2020). Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. En *ACS Combinatorial Science*. <https://doi.org/10.1021/acscombsci.0c00058>

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3). <https://doi.org/10.1093/sysbio/syq010>

Hadjithomas, M., & Moudrianakis, E. N. (2011). Experimental evidence for the role of domain swapping in the evolution of the histone fold. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1108649108>

Hanson-smith, V., Kolaczowski, B., & Thornton, J. W. (2010). *Robustness of Ancestral Sequence Reconstruction to Phylogenetic Uncertainty*. 27(9), 1988–1999. <https://doi.org/10.1093/molbev/msq081>

Hirota, S., Hattori, Y., Nagao, S., Taketa, M., Komori, H., Kamikubo, H., Wang, Z., Takahashi, I., Negi, S., Sugiura, Y., Kataoka, M., & Higuchi, Y. (2010). Cytochrome c polymerization by successive domain swapping at the C-terminal helix. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1001839107>

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., Grubmüller, H., & MacKerell, A. D. (2016). CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods*. <https://doi.org/10.1038/nmeth.4067>

Huang, Y., & Liu, Z. (2009). Kinetic Advantage of Intrinsically Disordered Proteins in

Coupled Folding-Binding Process: A Critical Assessment of the “Fly-Casting” Mechanism. *Journal of Molecular Biology*, 393(5). <https://doi.org/10.1016/j.jmb.2009.09.010>

Hubbard, R. E., & Kamran Haider, M. (2010). Hydrogen Bonds in Proteins: Role and Strength. En *Encyclopedia of Life Sciences*. <https://doi.org/10.1002/9780470015902.a0003011.pub2>

Huehn, J., Polansky, J. K., & Hamann, A. (2009). Epigenetic control of FOXP3 expression: The key to a stable regulatory T-cell lineage? En *Nature Reviews Immunology* (Vol. 9, Número 2). <https://doi.org/10.1038/nri2474>

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)

Ichiye, T., & Karplus, M. (1991). Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Structure, Function, and Bioinformatics*, 11(3). <https://doi.org/10.1002/prot.340110305>

Jackson, B. C., Carpenter, C., Nebert, D. W., & Vasiliou, V. (2010). Update of human and mouse forkhead box (FOX) gene families. *Human genomics*, 4(5), 345–352. <https://doi.org/2611113G3H18HNN3> [pii]

Janin, J., Miller, S., & Chothia, C. (1988). Surface, subunit interfaces and interior of oligomeric proteins. *Journal of Molecular Biology*. [https://doi.org/10.1016/0022-2836\(88\)90606-7](https://doi.org/10.1016/0022-2836(88)90606-7)

Janowski, R., Kozak, M., Abrahamson, M., Grubb, A., & Jaskolski, M. (2005). 3D domain-swapped human cystatin C with amyloidlike intermolecular  $\beta$ -sheets. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.20633>

Jerala, R., & Zerovnik, E. (1999). Accessing the global minimum conformation of stefin A dimer by annealing under partially denaturing conditions. *Journal of Molecular Biology*, 291(5), 1079–1089. <https://doi.org/10.1006/jmbi.1999.3045>

Kabsch, W., T., B. A., K., D., A., K. P., K., D., S., M., G., R. R. B., P., E., S., F., K., W., W., K., W., K., W., K., W., K., P., K., & S., W. M. (2010). XDS. *Acta Crystallographica Section D Biological Crystallography*, 66(2).

- Kang, X., Zhong, N., Zou, P., Zhang, S., Jin, C., & Xia, B. (2012). Foldon unfolding mediates the interconversion between M(pro)-C monomer and 3D domain-swapped dimer. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(37), 14900–14905. <https://doi.org/10.1073/pnas.1205241109>
- Kim, J. -S., & Raines, R. T. (1993). Ribonuclease S-peptide as a carrier in fusion proteins. *Protein Science*, *2*(3). <https://doi.org/10.1002/pro.5560020307>
- Kleywegt, G. J., & Brünger, A. T. (1996). Checking your imagination: Applications of the free R value. *Structure*. [https://doi.org/10.1016/S0969-2126\(96\)00097-4](https://doi.org/10.1016/S0969-2126(96)00097-4)
- Kumar, Sandeep, & Bansal, M. (1998). Dissecting  $\alpha$ -helices: Position-specific analysis of  $\alpha$ -helices in globular proteins. *Proteins: Structure, Function and Genetics*. [https://doi.org/10.1002/\(SICI\)1097-0134\(19980601\)31:4<460::AID-PROT12>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1097-0134(19980601)31:4<460::AID-PROT12>3.0.CO;2-D)
- Kumar, Shankar, Rosenberg, J. M., Bouzida, D., Swendsen, R. H., & Kollman, P. A. (1992a). THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, *13*(8), 1011–1021. <https://doi.org/10.1002/jcc.540130812>
- Kumar, Shankar, Rosenberg, J. M., Bouzida, D., Swendsen, R. H., & Kollman, P. A. (1992b). THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*. <https://doi.org/10.1002/jcc.540130812>
- Lalmansingh, A. S., Karmakar, S., Jin, Y., & Nagaich, A. K. (2012). Multiple modes of chromatin remodeling by Forkhead box proteins. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1819*(7), 707–715. <https://doi.org/10.1016/j.bbagr.2012.02.018>
- Larsen, T. A., Olson, A. J., & Goodsell, D. S. (1998). Morphology of protein–protein interfaces. *Structure*, *6*(4), 421–427. [https://doi.org/10.1016/S0969-2126\(98\)00044-6](https://doi.org/10.1016/S0969-2126(98)00044-6)
- Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, *25*(7). <https://doi.org/10.1093/molbev/msn067>
- Levy, Y., Onuchic, J. N., & Wolynes, P. G. (2007). Fly-casting in protein-DNA binding: Frustration between protein folding and electrostatics facilitates target recognition. *Journal of the American Chemical Society*, *129*(4). <https://doi.org/10.1021/ja065531n>

- Li, J., Dantas Machado, A. C., Guo, M., Sagendorf, J. M., Zhou, Z., Jiang, L., Chen, X., Wu, D., Qu, L., Chen, Z., Chen, L., Rohs, R., & Chen, Y. (2017). Structure of the Forkhead Domain of FOXA2 Bound to a Complete DNA Consensus Site. *Biochemistry*, *56*(29). <https://doi.org/10.1021/acs.biochem.7b00211>
- Liao, S. M., Du, Q. S., Meng, J. Z., Pang, Z. W., & Huang, R. B. (2013). The multiple roles of histidine in protein interactions. *Chemistry Central Journal*. <https://doi.org/10.1186/1752-153X-7-44>
- Littler, D. R., Alvarez-Fernández, M., Stein, A., Hibbert, R. G., Heidebrecht, T., Aloy, P., Medema, R. H., & Perrakis, A. (2010). Structure of the FoxM1 DNA-recognition domain bound to a promoter sequence. *Nucleic Acids Research*, *38*(13), 4527–4538. <https://doi.org/10.1093/nar/gkq194>
- Liu, L., Byeon, I. J. L., Bahar, I., & Gronenborn, A. M. (2012). Domain swapping proceeds via complete unfolding: A <sup>19</sup>F- and <sup>1</sup>H-NMR study of the cyanovirin-N protein. *Journal of the American Chemical Society*, *134*(9), 4229–4235. <https://doi.org/10.1021/ja210118w>
- Löytynoja, A. (2014). Phylogeny-aware alignment with PRANK. *Methods in Molecular Biology*. [https://doi.org/10.1007/978-1-62703-646-7\\_10](https://doi.org/10.1007/978-1-62703-646-7_10)
- Lu, M. M., Li, S., Yang, H., & Morrisey, E. E. (2002). Foxp4: A novel member of the Foxp subfamily of winged-helix genes co-expressed with Foxp1 and Foxp2 in pulmonary and gut tissues. *Mechanisms of Development*. [https://doi.org/10.1016/S0925-4773\(03\)00116-3](https://doi.org/10.1016/S0925-4773(03)00116-3)
- Ma, B., Elkayam, T., Wolfson, H., & Nussinov, R. (2003). Protein-protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(10). <https://doi.org/10.1073/pnas.1030237100>
- Ma, B., & Nussinov, R. (2007). Trp/Met/Phe Hot Spots in Protein-Protein Interactions: Potential Targets in Drug Design. *Current Topics in Medicinal Chemistry*, *7*(10). <https://doi.org/10.2174/156802607780906717>
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters

from ff99SB. En *Journal of Chemical Theory and Computation* (Vol. 11, pp. 3696–3713).  
<https://doi.org/10.1021/acs.jctc.5b00255>

Marianayagam, N. J., Sunde, M., & Matthews, J. M. (2004). The power of two: Protein dimerization in biology. En *Trends in Biochemical Sciences* (Vol. 29, Número 11).  
<https://doi.org/10.1016/j.tibs.2004.09.006>

Marsh, J. a., & Teichmann, S. a. (2015). Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annual Review of Biochemistry*, 84(1), 551–575.  
<https://doi.org/10.1146/annurev-biochem-060614-034142>

Medina, E., Córdova, C., Villalobos, P., Reyes, J., Komives, E. A., Ramírez-Sarmiento, C. A., & Babul, J. (2016). Three-Dimensional Domain Swapping Changes the Folding Mechanism of the Forkhead Domain of FoxP1. *Biophysical Journal*, 110(11), 2349–2360.  
<https://doi.org/10.1016/j.bpj.2016.04.043>

Medina, E., Villalobos, P., Coñuecar, R., Ramírez-Sarmiento, C. A., & Babul, J. (2019). The protonation state of an evolutionarily conserved histidine modulates domain swapping stability of FoxP1. *Scientific Reports*, 9(1), 5441. <https://doi.org/10.1038/s41598-019-41819-5>

Medina, E., Villalobos, P., Hamilton, G. L., Komives, E. A., Sanabria, H., Ramírez-Sarmiento, C. A., & Babul, J. (2020). Intrinsically Disordered Regions of the DNA-Binding Domain of Human FoxP1 Facilitate Domain Swapping. *Journal of Molecular Biology*.  
<https://doi.org/10.1016/j.jmb.2020.07.017>

Mendoza, E., Scharff, C., Rácz, I., Jarvis, E. D., & Wu, Y. (2017). *Protein-Protein Interaction Among the FoxP Family Members and their Regulation of Two Target Genes, VLDLR and CNTNAP2 in the Zebra Finch Song System*.  
<https://doi.org/10.3389/fnmol.2017.00112>

Mesquite Project Team. (2014). *Mesquite: A modular system for evolutionary analysis*. Available from: <http://mesquiteproject.wikispaces.com/home>.  
<https://doi.org/10.1017/CBO9781107415324.004>

Noel, J. K., Sulkowska, J. I., & Onuchic, J. N. (2010). Slipknotting upon native-like loop

- formation in a trefoil knot protein. *Proceedings of the National Academy of Sciences*, 107(35), 15403–15408. <https://doi.org/10.1073/pnas.1009522107>
- Noel, J. K., Whitford, P. C., & Onuchic, J. N. (2012). The shadow map: A general contact definition for capturing the dynamics of biomolecular folding and function. *Journal of Physical Chemistry B*, 116(29), 8692–8702. <https://doi.org/10.1021/jp300852d>
- Noel, J. K., Whitford, P. C., Sanbonmatsu, K. Y., & Onuchic, J. N. (2010). SMOG@ctbp: Simplified deployment of structure-based models in GROMACS. *Nucleic Acids Research*, 38(SUPPL. 2). <https://doi.org/10.1093/nar/gkq498>
- Nooren, I. M. A., & Thornton, J. M. (2003). *Structural Characterisation and Functional Significance of Transient Protein – Protein Interactions*. 2836(02), 991–1018. [https://doi.org/10.1016/S0022-2836\(02\)01281-0](https://doi.org/10.1016/S0022-2836(02)01281-0)
- Park, C., & Raines, R. T. (2000). Dimer formation by a “monomeric” protein. *Protein Science*. <https://doi.org/10.1110/ps.9.10.2026>
- Parker, M. J., Dempsey, C. E., Hosszu, L. L. P., Waltho, J. P., & Clarke, A. R. (1998). Topology, sequence evolution and folding dynamics of an immunoglobulin domain. *Nature Structural Biology*. <https://doi.org/10.1038/nsb0398-194>
- Perumal, K., Dirr, H. W., & Fanucchi, S. (2015). A Single Amino Acid in the Hinge Loop Region of the FOXP Forkhead Domain is Significant for Dimerisation. *The Protein Journal*, 34(2), 111–121. <https://doi.org/10.1007/s10930-015-9603-4>
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of computational chemistry*, 26(16), 1781–1802.
- Phillips, James C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *En Journal of Computational Chemistry*. <https://doi.org/10.1002/jcc.20289>
- Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P. R., Jenkins, H. T., Keegan, R., Krissinel, E., Stevenson, K., Lebedev, A., McNicholas, S. J., Nicholls, R. A., Noble, M., Pannu, N. S., Roth, C., Sheldrick, G., Skubak, P., Turkenburg, J., ... Wojdyr, M.

(2018). CCP 4 i 2: The new graphical user interface to the CCP 4 program suite. *Acta Crystallographica Section D: Structural Biology*, 74. <https://doi.org/10.1107/S2059798317016035>

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., Van Der Spoel, D., Hess, B., & Lindahl, E. (2013). GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29, 845–854. <https://doi.org/10.1093/bioinformatics/btt055>

Rajakulendran, T., Sahmi, M., Lefrançois, M., Sicheri, F., & Therrien, M. (2009). A dimerization-dependent mechanism drives RAF catalytic activation. *Nature*. <https://doi.org/10.1038/nature08314>

Ramírez-Sarmiento, C. A., Noel, J. K., Valenzuela, S. L., & Artsimovitch, I. (2015). Interdomain Contacts Control Native State Switching of RfaH on a Dual-Funneled Landscape. *PLoS Computational Biology*, 11(7), e1004379. <https://doi.org/10.1371/journal.pcbi.1004379>

Rogers, J. M., Waters, C. T., Seegar, T. C. M., Jarrett, S. M., Hallworth, A. N., Blacklow, S. C., & Bulyk, M. L. (2019). Bispecific Forkhead Transcription Factor FoxN3 Recognizes Two Distinct Motifs with Different DNA Shapes. *Molecular Cell*, 74(2). <https://doi.org/10.1016/j.molcel.2019.01.019>

Rousseau, F., Schymkowitz, J. W. H., Wilkinson, H. R., & Itzhaki, L. S. (2001). Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. *Proceedings of the National Academy of Sciences*, 98(10), 5596–5601. <https://doi.org/10.1073/pnas.101542098>

Sacco, M. D., Ma, C., Lagarias, P., Gao, A., Townsend, J. A., Meng, X., Dube, P., Zhang, X., Hu, Y., Kitamura, N., Hurst, B., Tarbet, B., Marty, M. T., Kolocouris, A., Xiang, Y., Chen, Y., & Wang, J. (2020). Structure and inhibition of the SARS-CoV-2 main protease reveal strategy for developing dual inhibitors against Mpro and cathepsin L. *Science Advances*. <https://doi.org/10.1126/sciadv.abe0751>

Schymkowitz, J. W. H., Rousseau, F., Wilkinson, H. R., Friedler, A., & Itzhaki, L. S. (2001). Observation of signal transduction in three-dimensional domain swapping. *Nature Structural*

*Biology*. <https://doi.org/10.1038/nsb1001-888>

Schymkowitz, J. W., Rousseau, F., Wilkinson, H. R., Friedler, a, & Itzhaki, L. S. (2001). Observation of signal transduction in three-dimensional domain swapping. *Nature structural biology*, 8(10), 888–892. <https://doi.org/10.1038/nsb1001-888>

Seeliger, M. A., Schymkowitz, J. W. H., Rousseau, F., Wilkinson, H. R., & Itzhaki, L. S. (2002). Folding and association of the human cell cycle regulatory proteins ckshs1 and ckshs2. *Biochemistry*. <https://doi.org/10.1021/bi0113465>

Staniforth, R. A., Giannini, S., Higgins, L. D., Conroy, M. J., Hounslow, A. M., Jerala, R., Craven, C. J., & Waltho, J. P. (2001). Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily. *EMBO Journal*, 20(17), 4774–4781. <https://doi.org/10.1093/emboj/20.17.4774>

Stroud, J. C., Wu, Y., Bates, D. L., Han, A., Nowick, K., Paabo, S., Tong, H., & Chen, L. (2006). Structure of the forkhead domain of FOXP2 bound to DNA. *Structure*, 14(1), 159–166. <https://doi.org/10.1016/j.str.2005.10.005>

Teufel, A., Wong, E. A., Mukhopadhyay, M., Malik, N., & Westphal, H. (2003). FoxP4, a novel forkhead transcription factor. *Biochimica et Biophysica Acta - Gene Structure and Expression*. [https://doi.org/10.1016/S0167-4781\(03\)00074-5](https://doi.org/10.1016/S0167-4781(03)00074-5)

Tsai, K. L., Sun, Y. J., Huang, C. Y., Yang, J. Y., Hung, M. C., & Hsiao, C. D. (2007). Crystal structure of the human FOXO3a-DBD/DNA complex suggests the effects of post-translational modification. *Nucleic Acids Research*, 35(20), 6984–6994. <https://doi.org/10.1093/nar/gkm703>

Vagin, A., & Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallographica Section D: Biological Crystallography*, 66(1). <https://doi.org/10.1107/S09074444909042589>

Valadares, N. F., De Oliveira-Silva, R., Cavini, I. A., De Almeida Marques, I., D’Muniz Pereira, H., Soares-Costa, A., Henrique-Silva, F., Kalbitzer, H. R., Munte, C. E., & Garratt, R. C. (2013). X-ray crystallography and NMR studies of domain-swapped canecystatin-1. *FEBS Journal*. <https://doi.org/10.1111/febs.12095>

- Webb, B., & Sali, A. (2014). Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics*, 47, 5 6 1-5 6 32. <https://doi.org/10.1002/0471250953.bi0506s47>
- Wu, Y., Borde, M., Heissmeyer, V., Feuerer, M., Lapan, A. D., Stroud, J. C., Bates, D. L., Guo, L., Han, A., Ziegler, S. F., Mathis, D., Benoist, C., Chen, L., & Rao, A. (2006). FOXP3 Controls Regulatory T Cell Function through Cooperation with NFAT. *Cell*, 126(2), 375–387. <https://doi.org/10.1016/j.cell.2006.05.042>
- Yang, S., Samuel, C., Levy, Y., Cheung, M. S., Levine, H., Wolynes, P. G., & Onuchic, J. N. (2004). Domain swapping is a consequence of minimal frustration. *Proceedings of the National Academy of Sciences of the United States of America*, 101(38), 13786–13791. <https://doi.org/10.1073/pnas.0403724101>
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8). <https://doi.org/10.1093/molbev/msm088>
- Yanshun, & Eisenberg. (2002). 3D domain swapping: As domains continue to swap. *Protein Science*, 11(6). <https://doi.org/10.1110/ps.0201402>
- Zhong, N., Zhang, S., Xue, F., Kang, X., Zou, P., Chen, J., Liang, C., Rao, Z., Jin, C., Lou, Z., & Xia, B. (2009). C-terminal domain of SARS-CoV main protease can form a 3D domain-swapped dimer. *Protein Science*. <https://doi.org/10.1002/pro.76>
- Zhu, R., Song, Y., Liu, H., Yang, Y., Wang, S., Yi, C., & Chen, P. R. (2017). Allosteric histidine switch for regulation of intracellular zinc(II) fluctuation. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1708563115>