
UNIVERSIDAD NACIONAL DE ROSARIO

**INSTITUTO DE BIOLOGÍA MOLECULAR Y CELULAR DE ROSARIO
(CONICET-UNR)**

*“Bases estructurales del reconocimiento ARN-proteína en el
procesamiento de pequeños ARNs en plantas”*

Tesis para optar título de Doctor en Ciencias Biológicas

Lic Paula Burdisso

Director: Dr. Rodolfo Rasia

2013

AGRADECIMIENTOS

Al Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET) y a la Agencia Nacional de Promoción Científica y Tecnológica por las becas otorgadas para la realización de esta Tesis.

A la Facultad de Ciencias Bioquímicas y Farmacéuticas de la UNR y al Instituto de Biología Molecular y Celular de Rosario (IBR) por facilitarme el equipamiento y el lugar de trabajo.

Al Dr. Rodolfo Rasia por permitirme formar parte de su grupo, por darme la posibilidad de crecer en el ámbito del laboratorio y principalmente, por estar presente en todo momento en que necesité su ayuda. Gracias por darme la posibilidad de expresar mi opinión libremente y confiar en mi para llevar a cabo este proyecto.

Al Dr. Alejandro Vila, por permitirnos compartir el laboratorio y facilitarnos el material para realizar nuestro trabajo.

Al Dr. Oscar Roveri y a todos los miembros de la Cátedra de Biofísica, por todos estos años compartidos y por sus contribuciones a mi formación.

A mis tutores de Tesis, los Dres. Jorgelina Ottado y Claudio Fernandez, por su buena predisposición y críticas constructivas.

A todos nuestros colaboradores, especialmente al grupo del Dr. Javier Palatnik que me brindaron las herramientas y el conocimiento necesario para poder entrar en *"el mundo de Arabidopsis"*.

A todos mis compañeros de laboratorio, los de suipacha y los del CCT. Gracias por la buena onda. Especialmente quiero agradecer a Irina, Marcos, Fernando y Guillermo, con quienes compartí la mesada codo a codo. Gracias por los buenos momentos.

A mis amigos del alma, gracias por brindarme su amistad y acompañarme en las buenas y en las malas.

A mi familia, por ser mi sostén y brindarme todo el apoyo que necesito. Gracias por no abandonarme en ningún momento y por estar siempre a mi lado.

A Gustavo, por ser incondicional. Gracias por la paciencia y el amor.

A mi abuelo, que me ofreció su corazón desde que comencé a dar mis primeros pasos. Gracias Toto por tu divertida e inigualable compañía de la que jamás voy a dejar de valorar.

Parte de los trabajos realizados en esta tesis han sido presentados en las siguientes publicaciones:

Rasia, R.M., Mateos J.L., Bologna N.G., **Burdisso P.**, Imbert L., Palatnik J. F. and Boisbouvier J., (2010) "*Structure and RNA interactions of the plant microRNA processing associated protein HYL1*" *Biochemistry*. (38):8237-8239.

Burdisso, P., Suarez, I. P., Bologna, N. G., Palatnik, J. F., Bersch, B., and Rasia, R. M., (2012) "*Second dsRBD of DCL1: Structure and nucleic acid binding*", *Biochemistry*, (51): 10159–10166.

ÍNDICE.....	i
ABREVIATURAS.....	v
RESUMEN.....	vii
1 INTRODUCCIÓN.....	1
1.1 SILENCIAMIENTO GÉNICO.....	1
1.1.1 DESCUBRIMIENTO DEL ARN DE INTERFERENCIA.....	1
1.1.2 SILENCIAMIENTO GÉNICO POR ARN PEQUEÑOS EN PLANTAS.....	2
1.1.3 ARN PEQUEÑOS DE PLANTAS.....	5
1.1.3.1 siARNs, ta-siARNs y nat-ARNs.....	5
1.1.3.2 miARNs.....	8
1.2 PROCESAMIENTO DE miARNs.....	9
1.2.1 BIOGÉNESIS DE miARNs EN ANIMALES.....	10
1.2.2 BIOGÉNESIS DE miARNs EN PLANTAS.....	11
1.2.2.1 PROTEÍNAS DE PROCESAMIENTO DE PLANTAS.....	12
1.2.2.1.1 DICER.....	13
1.2.2.1.2 HYL1.....	15
1.2.2.1.3 SERRATE.....	17
1.2.2.1.4 CBC.....	18
1.3 REGULACIÓN DE LA VÍA DE miARNs.....	18
1.4 RECONOCIMIENTO DE SUSTRATO EN LA BIOGÉNESIS DE miARNs.....	20
1.4.1 DETERMINANTES ESTRUCTURALES DE LOS PRECURSORES DE miARN.....	20
1.5 DOMINIOS dsRBDs.....	22
1.5.1 CARACTERÍSTICAS ESTRUCTURALES DE LOS dsRBDs.....	27
1.5.2 VARIACIONES RESPECTO A LOS dsRBDs CANÓNICOS.....	29
1.5.3 CARACTERÍSTICAS ESTRUCTURALES DEL SUSTRATO DE dsRBDs.....	31
1.5.4 DESCRIPCIÓN DE LA INTEFASE DE UNIÓN dsRBD-ARN.....	33
1.5.4.1 RECONOCIMIENTO DEL SURCO MAYOR POR LA REGIÓN 3.....	33
1.5.4.2 RECONOCIMIENTO DEL SURCO MENOR DEL ARNdh.....	34
1.6 FUNCIONES DE LOS dsRBDs EN LA BIOGÉNESIS DE miARNs.....	36
1.6.1 dsRBDs DE DCL1.....	37
1.6.2 dsRBDs DE HYL1.....	39
1.6.3 RECONOCIMIENTO DEL SITIO DE CORTE DE LOS PRECURSORES POR dsRBDs.....	41

2 OBJETIVOS.....	43
2.1 OBJETIVOS GENERALES.....	43
2.2 OBJETIVOS ESPECÍFICOS.....	43
3 MATERIALES Y MÉTODOS.....	44
3.1 CEPAS, PLÁSMIDOS Y MEDIOS DE CULTIVO.....	44
3.1.1 CEPAS BACTERIANAS.....	44
3.1.2 VECTORES PLASMÍDICOS.....	44
3.1.3 MEDIOS DE CULTIVO.....	46
3.2 TRANSFORMACIÓN BACTERIANA.....	46
3.2.1 PRODUCCIÓN DE CÉLULAS COMPETENTES.....	46
3.2.2 TRANSFORMACIÓN DE E.coli.....	47
3.2.3 TRANSFORMACIÓN DE A. tumefaciens.....	47
3.2.4 SELECCIÓN Y ANÁLISIS DE TRANSFORMANTES.....	48
3.3 SUBCLONADO DE FRAGMENTOS DE ADN.....	48
3.3.1 MINIPREPARACIONES DE ADN PLASMÍDICO.....	48
3.3.2 ESTIMACIÓN DE LA CONCENTRACIÓN DE ADN.....	49
3.3.3 AMPLIFICACIÓN DE FRAGMENTOS DE ADN MEDIANTE REACCIÓN EN CADENA DE LA POLIMERASA (PCR).....	49
3.3.4 ELECTROFORESIS DE ADN EN GELES DE AGAROSA.....	50
3.3.5 PURIFICACIÓN DE FRAGMENTOS DE ADN A PARTIR DE GELES DE AGAROSA.....	50
3.3.6 DIGESTIÓN DE ADN CON ENDONUCLEASAS DE RESTRICCIÓN.....	51
3.3.7 LIGACIÓN DE FRAGMENTOS DE ADN.....	51
3.3.8 CHEQUEO DE TRANSFORMANTES.....	51
3.3.8.1 PCR DE COLONIAS.....	51
3.3.8.2 ANÁLISIS DE RESTRICCIÓN.....	52
3.3.8.3 SECUENCIACIÓN DE ADN.....	52
3.4 EXPRESIÓN Y PURIFICACIÓN DE PROTEÍNAS.....	52
3.4.1 EXPRESIÓN DE PROTEÍNAS RECOMBINANTES.....	52
3.4.2 PURIFICACIÓN DE PROTEÍNAS EN CONDICIONES NATIVAS.....	53
3.4.3 PURIFICACIÓN DE PROTEÍNAS EN CONDICIONES DESNATURALIZANTES.....	54
3.4.4 REPLEGADO DE PROTEÍNAS.....	55
3.4.5 PURIFICACIÓN DE LA PROTEASA TEV.....	55

3.4.6 ESTIMACIÓN DE LA CONCENTRACIÓN DE PROTEÍNAS.....	56
3.4.7 ELECTROFORESIS DE PROTEÍNAS EN GELES DE POLIACRILAMIDA.....	56
3.5 ESPECTROSCOPÍA DE RESONANCIA MAGNÉTICA NUCLEAR.....	57
3.5.1 FUNDAMENTOS TEÓRICOS.....	57
3.5.1.1 EL ESPECTRO 1H15N-HSQC.....	57
3.5.1.2 ASIGNACIÓN DE RESONANCIAS 1H, 15N Y 13C.....	58
3.5.1.3 CÁLCULO DE LA ESTRUCTURA EN SOLUCIÓN.....	59
3.5.1.4 ACOPLAMIENTOS RESIDUALES DIPOLARES.....	62
3.5.1.5 DETERMINACIÓN DE LOS PARÁMETROS DINÁMICOS EN SOLUCIÓN.....	63
3.5.1.5.1 PARÁMETROS DE RELAJACIÓN.....	63
3.5.1.5.2 DETERMINACIÓN DEL TENSOR DE DIFUSIÓN ROTACIONAL.....	67
3.5.1.5.3 ANÁLISIS MODEL FREE.....	68
3.5.1.6 INTERCAMBIO QUÍMICO.....	71
3.5.2 PROCEDIMIENTOS EXPERIMENTALES.....	73
3.5.2.1 ASIGNACIÓN DE RESONANCIAS DE DCL1-B.....	73
3.5.2.2 DETERMINACIÓN DE RDCs DE DCL1-B.....	74
3.5.2.3 CÁLCULO DE ESTRUCTURA EN SOLUCIÓN POR RMN.....	74
3.5.2.4 INTERACCIÓN PROTEÍNA-PROTEÍNA Y PROTEÍNA-ÁCIDOS NUCLEICOS.....	75
3.5.2.5 DINÁMICA DE DCL1-B.....	76
3.6 SÍNTESIS DE ARN.....	76
3.7 ESPECTRO DE EMISIÓN DE FLUORESCENCIA.....	77
3.8 TITULACIONES POR ANISOTROPÍA DE LA FLUORESCENCIA.....	77
3.9 DICROÍSMO CIRCULAR.....	79
3.10 OBTENCIÓN DE PLANTAS DE <i>A. thaliana</i> TRANSFORMADAS.....	80
3.10.1 ECOTIPO UTILIZADO Y CONDICIONES DE CRECIMIENTO.....	80
3.10.2 TRANSFORMACIÓN DE <i>A.thaliana</i>	80
3.10.3 RECOLECCIÓN DE SEMILLAS.....	81
3.10.4 ESTERILIZACIÓN DE SEMILLAS.....	81
3.10.5 SELECCIÓN DE TRANSFORMANTES.....	81
3.11 MUESTRAS VEGETALES.....	82
3.11.1 RECOLECCIÓN DE MUESTRAS.....	82
3.11.2 EXTRACCIÓN DE PROTEÍNAS DE INFLORESCENCIAS.....	82

3.11.3 CUANTIFICACIÓN DE PROTEÍNAS.....	82
3.12 WESTERN BLOT.....	83
3.13 MICROSCOPIA CONFOCAL.....	84
3.14 PROGRAMAS UTILIZADOS.....	84
4 CAPÍTULO 1: dsRBD2 de DCL1.....	86
4.1 RESULTADOS 1.....	86
4.1.1 RESULTADOS PREVIOS.....	86
4.1.2 OPTIMIZACIÓN DE CONDICIONES PARA EL ESTUDIO DE LA PROTEÍNA POR RMN....	88
4.1.2.1 EVALUACIÓN DE PROTOCOLOS DE EXPRESIÓN.....	89
4.1.2.2 EVALUACIÓN DE PROTOCOLOS DE PURIFICACIÓN.....	90
4.1.3 ESPECTROS DE FLUORESCENCIA Y DICROÍSMO CIRCULAR.....	92
4.1.4 ASIGNACIÓN DE LAS RESONANCIAS DE DCL1-B Y CSI.....	94
4.1.5 CÁLCULO DE ESTRUCTURA EN SOLUCIÓN.....	97
4.1.5.1 CÁLCULO DE ESTRUCTURA POR RDC.....	102
4.1.6 INTERACCIÓN PROTEÍNA-PROTEÍNA.....	106
4.1.7 UNIÓN DE DCL1-B A ÁCIDOS NUCLEICOS.....	107
4.1.8 MEDIDAS DE DINÁMICA POR RESONANCIA MAGNÉTICA NUCLEAR.....	113
4.1.8.1 ANÁLISIS DE LOS DATOS DE RELAJACIÓN.....	113
4.1.8.2 ANÁLISIS "MODEL FREE".....	114
4.1.9 LOCALIZACIÓN DE DCL1-B IN VIVO.....	116
4.2 DISCUSIÓN 1.....	119
5 CAPÍTULO 2: dsRBD1 de HYL1.....	123
5.1 RESULTADOS 2.....	123
5.1.1 RESULTADOS PREVIOS.....	123
5.1.2 FUNCIÓN IN VIVO DE LAS VARIANTES DE HYL1-A.....	124
5.1.3 ANÁLISIS ESTRUCTURAL DE LOS DOMINIOS MUTANTES.....	128
5.1.4 ESTABILIDAD Y UNIÓN AL SUSTRATO.....	130
5.2 DISCUSIÓN 2.....	133
6 CONCLUSIONES GENERALES.....	137
7 BIBLIOGRAFÍA.....	140

ABREVIATURAS

ADN	Ácido desoxirribonucleico.
ADNasa	ADN nucleasa.
ADNc	ADN complementario.
AGO	Proteína de la familia ARGONAUTA.
ARN	Ácido ribonucleico.
ARNasa	ARN nucleasa.
ARNdh	ARN doble hebra.
ARNi	ARN de interferencia.
ARNm	ARN mensajero.
ARNsh	ARN simple hebra.
CBC	Complejo de unión a cap.
Col	Ecotipo Columbia-0 de <i>Arabidopsis thaliana</i> .
DCL	Ribonucleasa Dicer – Like.
DCR	Dicer
DGCR8	Di George syndrome critical region gene 8.
dsRBD	Dominios de unión a ARNdh.
EXP5	Exportin-5.
HEN1	Hua enhancer 1.
HSQC	Experimento de coherencia cuántica simple heteronuclear.
HYL1	Hyponastic leaves 1.
LB	Luria-Bertani.
M9	Medio mínimo M9
MiARN	MicroARN.
Min	Minuto/s.
MS	Murashige-Skoog.
nat-miARN	miARNs naturales.
nat-siARN	siARNs naturales.
NLS	Señal de localización nuclear

Nt	Nucleótidos.
Pb	Pares de base.
PCR	Reacción en cadena de la polimerasa.
PKR	RNA dependent protein kinase
Pol II	ARN polimerasa II.
RDR	ARN polimerasas dependientes de ARN.
RISC	Complejo de silenciamiento inducido por ARN.
RMN	Resonancia Magnética Nuclear.
RT	Retrotranscripción.
RT-qPCR	Transcripción reversa seguida de PCR en tiempo real cuantitativa.
SE	Serrate.
seg	Segundo/s.
siARN	ARN pequeño interferente.
ta-siARN	trans-acting siARNs.

RESUMEN

Los microARNs (miARNs) son moléculas de ARN pequeñas de 21 nucleótidos de longitud que se sintetizan en el núcleo por la ARN polimerasa II. En plantas, están involucrados en la regulación de procesos como el desarrollo, resistencia a estrés y respuestas a hormonas. La biogénesis de miARNs comienza con la transcripción de precursores más largos, con extensa estructura secundaria de tallo y burbuja dentro de los cuales está contenida la secuencia que corresponde al mensaje de 21 nucleótidos. Estos precursores son procesados por un complejo proteico formado por la ARNasa III DICER-LIKE 1 (DCL1) y las proteínas accesorias HYPONASTIC LEAVES 1 (HYL1) y SERRATE (SE). Los precursores de plantas son sumamente heterogéneos. Sin embargo, la maquinaria de procesamiento, es capaz de liberar con precisión el miARN que posteriormente efectuará su acción regulando negativamente ARN mensajeros por complementariedad de bases de Watson y Crick. Durante los últimos años, se han dedicado muchos esfuerzos en descubrir nuevos miARNs, así como también se han logrado numerosos avances respecto a las formas de procesamiento de los precursores de plantas. Sin embargo, el mecanismo por el cual las proteínas de procesamiento llevan a cabo el reconocimiento es hasta el momento poco conocido. En este trabajo realizamos una caracterización biofísica de los dominios de unión a ARN doble hebra. En primer lugar se calculó la estructura en solución del segundo dsRBD de DCL1, a partir de la cual se observó que si bien tiene un plegamiento de dsRBD canónico, presenta diferencias con respecto a dominios homólogos. También se demostró que este dominio es capaz de unir tanto ARNdh como ADN, en contraste con lo que ocurre con la mayoría de los dsRBDs. La caracterización funcional de este dominio demostró que posiblemente actúe como una señal de localización atípica para direccionar a DCL1 al núcleo. Por otro lado, se analizaron los distintos determinantes de unión a sustrato del primer

dsRBD de la proteína accesoria HYL1. Para esto, se generaron formas mutantes de la proteína, las cuales mantienen su estructura global, pero afectan las propiedades de unión al sustrato. Sorprendentemente, se demostró que una mutación y hasta una delección completa en la región que se propone como principal determinante de unión al ARN no causa mayores efectos. Finalmente, se analizó la función de cada mutante *in vivo*, estableciendo una correlación directa entre la afinidad por los precursores y la actividad de la proteína.

1 INTRODUCCIÓN

1.1 SILENCIAMIENTO GÉNICO

1.1.1 DESCUBRIMIENTO DEL ARN DE INTERFERENCIA

El ARN de interferencia (ARNi), es un proceso natural utilizado por las células para “apagar” o silenciar la actividad de genes específicos. Los primeros pasos en el descubrimiento del fenómeno del ARNi, fueron dados en 1990 por Napoli y Jorgensen ¹, durante la realización de un trabajo que estaba dedicado a investigar la ruta de pigmentos de flores de petunia. Con el objetivo de obtener una coloración violeta más intensa, estos investigadores sobreexpresaron en petunias silvestres la chalcona sintasa (CHS), una enzima clave en la síntesis de flavonoides. Sorpresivamente, luego de realizar dichos experimentos, hallaron que algunas de las flores obtenidas eran blancas y que los niveles de CHS en éstas eran 50 veces menores que en petunias salvajes. En 1992, Romano y Marcino reportaron un fenómeno similar en *Neurospora crassa* ². Notaron que la introducción de secuencias de ARN homólogas a un gen endógeno, causaban la represión del mismo. Los primeros reportes de silenciamiento génico en animales, se obtuvieron en el año 1995 en un trabajo realizado en *Caenorhabditis elegans* ³. En el mismo, Guo y Kempfues observaron que la introducción de ARN antisentido al ARN mensajero (ARNm) del gen *par-1*, resultaba en la degradación del transcripto correspondiente. De manera inesperada, cuando se realizaron experimentos control usando sólo el ARN sentido de *par-1*, que es incapaz de hibridar con el transcripto de *par-1* endógeno, observaron que el mensajero de este gen seguía siendo blanco para la degradación. En 1998, Fire y Mello publicaron un trabajo en el que demostraron que la inyección de ARN doble hebra (ARNdh), una mezcla de ARN sentido y ARN antisentido en *C. elegans*, era mucho más eficiente en el silenciamiento del gen,

que la inyección de ARN simple hebra (ARNsh). En el mismo trabajo, también demostraron que la inyección de unas pocas moléculas de ARNdh por célula era suficiente para silenciar la expresión del gen, y más interesante aún, fue el descubrimiento de que la inyección de ARNdh en el intestino del gusano, provocaba que el silenciamiento se dispare en todo organismo e incluso, se transmita a la primera generación de descendientes. Este trabajo, publicado en 1998 en la revista Nature, dio el puntapié inicial en el estudio y comprensión de la regulación de la expresión génica mediante ARN pequeños. En el año 2006 los autores del trabajo fueron reconocidos por la Academia Real Sueca de Ciencias, al recibir el Premio Nobel de Fisiología o Medicina 2006.

Al hallar que la inyección de ARNdh en un gusano o planta, causaba un silenciamiento sistémico y que además, era transmitido a la siguiente generación, los investigadores sospecharon que este mecanismo estaba mediado por alguna molécula efectora estable del silenciamiento. Los primeros estudios en búsqueda de estos intermediarios fueron realizados en 1999 por Hamilton y Baulcombe, mientras estudiaban el fenómeno de silenciamiento génico post-transcripcional inducido por virus en plantas ⁴. Si bien en ese momento existía la idea general de que el ARNdh tenía que separarse para que la hebra antisentido se una al ARNm, la hebra antisentido de longitud completa nunca fue detectada. Este hecho, llevó a Hamilton y Baulcombe a buscar formas más cortas del ARN antisentido derivado del ARNdh. La hipótesis de estos investigadores, consistía en que el ARN antisentido podría servir como una guía, uniéndose al ARNm complementario y causando su degradación. Al descubrir que el ARN antisentido tenía una longitud estimada de 25 nucleótidos (nt), sugirieron que esta cantidad de nt era la necesaria para la especificidad del ARNi. Al año siguiente, dos grupos independientes, identificaron que ARNs de 21-23 nucleótidos co-purificaban con el ARNi, sugiriendo que el ARNdh era convertido en intermediarios más cortos a

los que llamaron ARN pequeños de interferencia (siARN por sus siglas en inglés) y que eran capaces de unirse a ARNm complementarios llevando a la degradación del transcripto blanco ^{5,6}.

El descubrimiento de las moléculas de ARN pequeñas, con capacidad de regular genes e incluso genomas, ha sido uno de los avances más significativos en biología en la última década. Estas moléculas son ARN no codificantes de aproximadamente 20-30 nucleótidos de longitud que actúan como guías para la regulación de múltiples procesos biológicos. Los efectos de estos ARN pequeños sobre el control de genes es generalmente, inhibitorio y el mecanismo regulatorio correspondiente se incluye dentro de los mecanismos de silenciamiento génico.

1.1.2 SILENCIAMIENTO GÉNICO POR ARN PEQUEÑOS EN PLANTAS

Si bien existen muchas clases de ARN pequeños, varios aspectos de sus orígenes, estructuras, proteínas efectoras asociadas y funciones biológicas, han llevado a un reconocimiento de 3 categorías principales: los ARN pequeños de interferencia (siARNs), los micro ARNs (miARNs) y los ARN que interactúan con piwi (piARNs). Los siARN y miARN están ampliamente distribuidos en términos filogenéticos y fisiológicos, en contraste, los piARN se encuentran principalmente en animales, y tienen funciones fundamentalmente en la línea germinal. En todos los casos, la identidad de los genes que van a ser silenciados está especificada por la secuencia del ARN pequeño, que es capaz de reconocer a su blanco por complementariedad de bases de Watson y Crick.

En plantas existen distintos tipos de ARN pequeños. Si bien cada grupo posee características particulares en cuanto a su origen, tamaño y mecanismo de acción, el proceso de silenciamiento génico mediado por ARN pequeños comparte cuatro pasos conservados y necesarios: (i) la presencia de una

molécula de ARNdh, (ii) el procesamiento del ARNdh para liberar un ARN pequeño de un tamaño de 18-25 nt de longitud, (iii) la metilación del extremo 3' del ARN pequeño y (iv) por último, la incorporación del ARN pequeño simple hebra al complejo efector que contiene una proteína de la familia ARGONAUTA (AGO) ⁷.

El origen de la molécula de ARNdh que va a ser procesada es diverso. La misma, puede derivar de la replicación de un virus, de la transcripción de secuencias invertidas, de estructuras de tallo y bucle intramoleculares de un transcripto o de la síntesis de un ARNdh por acción de ARN polimerasas dependientes de ARN (RDR, del inglés, *RNA-dependent RNA polymerase*). El procesamiento del ARNdh involucra la participación de ribonucleasas de tipo III denominadas DICER. Las proteínas DICER se encuentran altamente conservadas en las células eucariotas y están presentes tanto en organismos unicelulares como en plantas superiores y vertebrados. En *Arabidopsis* existen cuatro proteínas homologas a DICER denominadas DICER-LIKE (DCL). DCL1 produce ARN pequeños entre 18-21 nt, mientras que DCL2, DCL3 y DCL4 producen ARN pequeños de un tamaño de 22 nt, 24 nt y 21 nt de longitud, respectivamente ⁸. Luego de que DCL produce la liberación del dúplex miARN/miARN* o siARN/siARN* esta molécula sufre una metilación por la proteína HEN1 e inmediatamente la hebra molde es incorporada al complejo de silenciamiento RISC (por sus siglas en inglés, *RNA induced silencing complex*) y lleva a cabo la búsqueda de ácidos nucleicos complementarios para ejecutar su función inhibitoria.

Una vez que el fragmento de ARN pequeño es incorporado en el complejo RISC, pueden tener lugar los siguientes eventos: (i) corte endonucleotídico sobre ARN mensajeros blancos, (ii) represión de la traducción de los mismos (silenciamiento post-transcripcional) o (iii) metilación de citosinas del ADN o de

histonas (silenciamiento transcripcional) ⁹. El tipo de función efectora que se lleve a cabo, va estar definida por la identidad de la proteína AGO que forme parte del complejo RISC.

En plantas, existe redundancia en los componentes de la maquinaria de ARNi. *A. thaliana* además de las cuatro DCLs, tiene diez proteínas AGO y seis polimerasas dependientes de ARN (RDRs). Todas ellas están altamente conservadas y colaboran en varias etapas dentro de las diferentes rutas de silenciamiento. El modo en que cada clase de siARN es dirigido hacia una única vía es hasta el momento poco conocido.

1.1.3 ARN PEQUEÑOS DE PLANTAS

Los ARN pequeños regulatorios de plantas pueden ser agrupados en cuatro grupos en función de las diferencias en su biogénesis y su modo de acción. La clase más abundante está integrada por los siARNs, mientras que el segundo lugar en abundancia está ocupado por los miARNs. Los otros dos grupos de ARN pequeños que han sido descubiertos posteriormente, están conformados por los ta-siARN (del inglés, trans-acting siARNs), y por los nat-siARNs y nat-miARNs (del inglés, *natural antisense siRNAs* y *miRNAs*) ^{7,10-12}

1.1.3.1 siARNs, ta-siARNs y nat-ARNs

Los **siARNs** representan a los inductores de ARNi canónicos. Sus precursores son moléculas de ARNd_h extensas, con complementariedad de bases perfecta y de estructura lineal (Figura 1). Dichos precursores, pueden ser introducidos directamente desde el medio ambiente o pueden ser producidos endógenamente por la célula. Los primeros siARNs identificados, se observaron durante el silenciamiento génico inducido por transposones y virus en plantas ¹³. Una vez que estas moléculas de ARNd_h son detectadas en el citoplasma de la célula, se induce la respuesta de defensa, donde el sustrato es procesado por

DCL2 o DCL4, generando siARNs de 21-22 nt que serán los efectores del silenciamiento post-transcripcional ¹⁴. En los años 2002 y 2003 se descubrió que existía otro tipo de siARNs endógenos. Sus precursores son generados a partir de ciertas regiones genómicas como centrómeros, regiones de inserción de transposones u otras secuencias repetitivas que se transcriben por la ARN polimerasa IVa. Una vez transcritas, el ARNsh es procesado por una RDR, formando ARNdsh que a continuación es procesado por DCL3 para liberar finalmente siARNs de 24 nt de longitud que llevan a cabo el silenciamiento a nivel transcripcional.

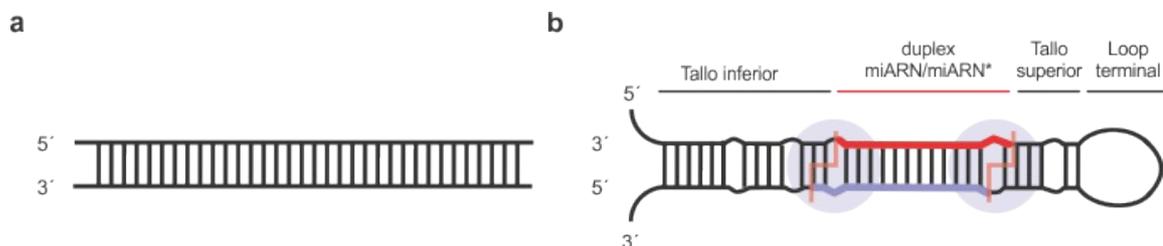


Figura 1: Esquema de los precursores de a) siARNs y b) miARNs

Los **siARN trans-acting** (ta-siARNs) se descubrieron a partir de estudios funcionales en plantas. La biogénesis de estas moléculas de ARN pequeñas es especialmente compleja (Figura 2). Los genes que dan lugar a los ta-siRNAs son transcritos por la ARN polimerasa II dando lugar a un ARN no codificante denominado TAS. Este ARN es reconocido por un miARN complementario, lo que provocará que AGO1 realice un corte de manera tal que el miARN pueda actuar como cebador para la transcripción de ARNdsh por parte de una RDR específica. Finalmente el ARNdsh generado es procesado por DCL4, que realiza cortes sucesivos secuenciales, liberando siARNs de 21 nts cuya secuencia esta especificada por la fase generada a partir del corte del miARN inicial. Algunos de los ARN pequeños formados, son incorporados al complejo RISC y guían la degradación de ARNm complementarios ¹⁵⁻¹⁷.

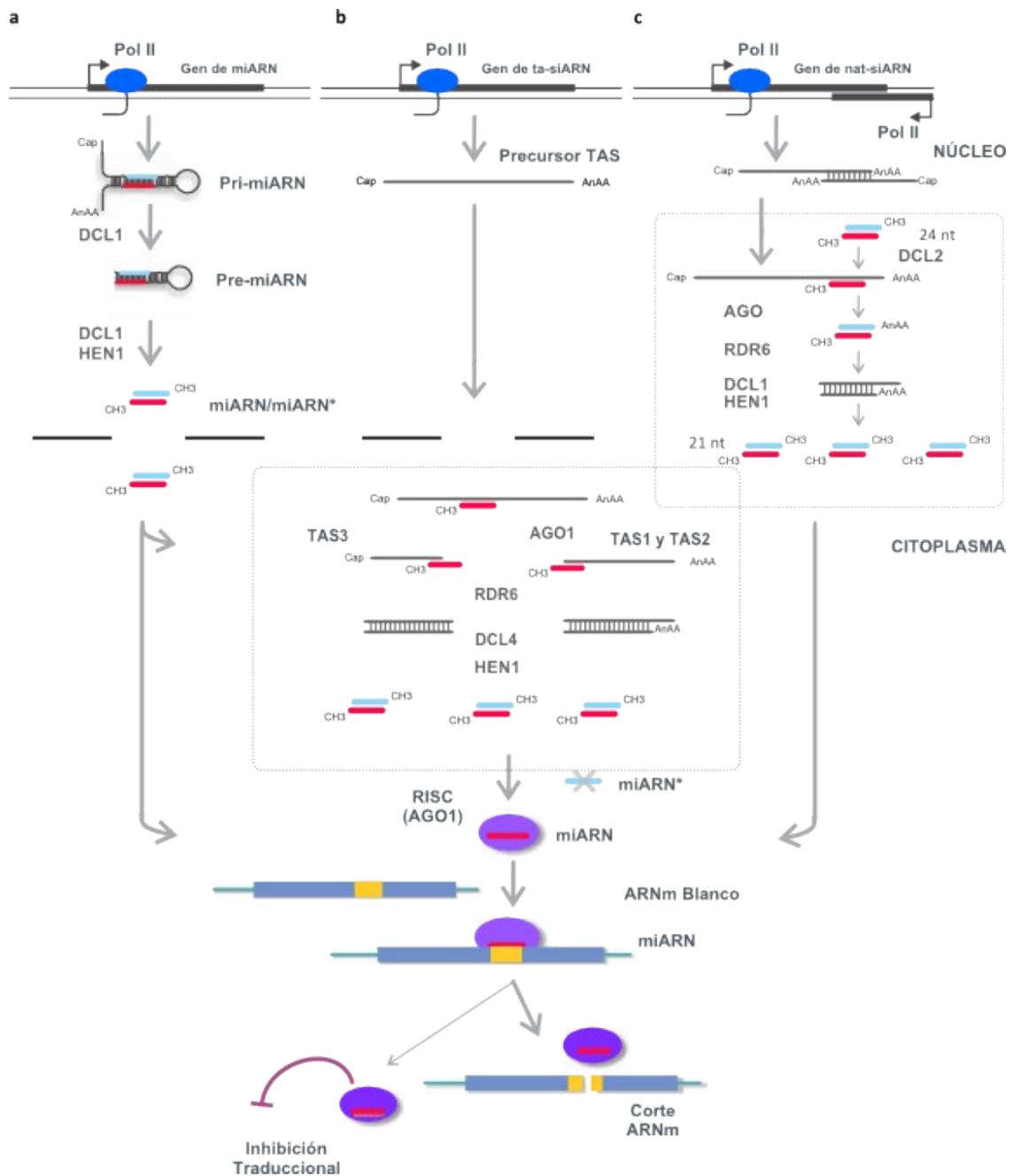


Figura 2: Vías de ARN pequeños de *A. thaliana* a) miARNs; b) ta-siARNs y c) nat-siARNs

El último grupo de moléculas de ARN pequeñas de plantas son los **siARN naturales antisentido** (nat-siARN), que se originan a partir de ARNdhs formados por la transcripción bidireccional convergente de dos genes parcialmente

solapados ^{18,19}. En los casos reportados hasta el momento siempre uno de los genes solapados presenta una expresión constitutiva. La inducción de la expresión del gen complementario activa la formación del ARNdh el cual es procesado por DCL1 o DCL2 liberando siARN primarios (Figura 2). Estos siARN primarios son capaces de inducir el corte sobre el ARNm expresado constitutivamente. A continuación, una RDR sintetiza la hebra complementaria del ARNm cortado. Este corte, genera una molécula de ARNdh que es procesada de manera procesiva por DCL1 para producir una sucesión de siARNs denominados nat-siARN secundarios y que están ubicados en fase de 21 nt cada uno ^{18,19}.

1.1.3.2 *miARNs*

Los **miARN** son moléculas de ARN de aproximadamente 21 nt de longitud que regulan negativamente la expresión de ARN mensajeros específicos a nivel post-transcripcional. Se procesan a partir de precursores más largos (pri-miARNs) que son transcritos en el núcleo por la ARN polimerasa II. A diferencia de los precursores de siARNs, los pri-miARN forman estructuras en forma de hebilla, con un tallo doble hebra y un bucle terminal que posteriormente debe ser procesado para obtener la molécula efectora de 21 nt de longitud (Figura 1).

En la actualidad, se considera que aproximadamente un tercio de los genes de animales están regulados por miARNs ²⁰. Durante este último tiempo, se han identificado cientos de miARNs en diversos animales y plantas, al principio mediante clonado, y más tarde a través de análisis computacionales de genomas completos. Al presente hay anotados ca. 21264 miARNs (miRBase, agosto de 2012). A pesar de estos avances en investigaciones relacionadas a miARNs, de los miles de miARNs identificados, sólo se conocen las funciones biológicas de unos pocos. En el caso de los miARN de plantas, se ha observado que generalmente

presentan un alto grado de complementariedad con el ARNm blanco y que su principal función es llevar a cabo la digestión del mismo ²¹. Sin embargo, en algunos casos, se ha reportado que estas moléculas cumplen funciones inhibiendo la síntesis de proteínas ^{22,23}. Una característica particular de los blancos de miARNs de plantas conocidos hasta el momento es que la mayoría son factores de transcripción, y que están principalmente involucrados en la regulación del desarrollo o de la diferenciación celular. En contraste, los ARNms blanco de los miARNs de animales, generalmente están relacionados con funciones mucho más diversas. Una característica destacable de estas moléculas es que tanto en animales como en plantas, una sola especie de miARN puede unirse a diferentes ARNms blanco y por el contrario, muchos miARNs distintos, pueden cooperar en el control de un mismo ARNm. Esta posibilidad, genera una red compleja de regulación que asegura el funcionamiento correcto de los distintos procesos biológicos ²⁴.

1.2 PROCESAMIENTO DE miARNs

Como se mencionó previamente, los miARNs son codificados por el propio genoma. El genoma de *Arabidopsis* contiene cientos de regiones que codifican para miARNs denominadas genes MIR. Los loci de los genes MIR de animales tienden a estar en "clusters" de miARNs que luego son transcritos como precursores policistronicos ²⁵. Por el contrario, los genes MIR de plantas generalmente no están agrupados y cada gen se transcribe a un precursor que da lugar a una única especie de miARN. Al igual que los genes codificantes de proteínas, los genes MIR se transcriben por la ARN polimerasa II y los transcritos primarios sufren un agregado de cap en el extremo 5' y una poliadenilación en el extremo 3' ²⁶. El miARN está contenido dentro de la estructura de hebilla del precursor, que debe ser procesado para liberar el producto final. A pesar de que los mecanismos de generación de miARNs comparten características comunes en

animales y plantas, es importante destacar que existen diferencias respecto a las propiedades de sus precursores y al mecanismo de procesamiento de los mismos, lo que podría estar relacionado con los distintos roles que cumplen estas moléculas en cada una de las especies.

1.2.1 BIOGÉNESIS DE miARNs EN ANIMALES

La biogénesis de miARN en animales ha sido extensamente caracterizada. La misma, se lleva a cabo en dos pasos separados en tiempo y espacio. La primer etapa del procesamiento se realiza en el núcleo, donde el complejo microprocesador formado principalmente por la ARNasa III DROSHA y la proteína de unión a ARNdH DGCR8/PASHA (por sus siglas en inglés, *DiGeorge síndrome critical región gene 8*), realiza el primer corte del pri-miARN liberando un precursor más corto, el pre-miARN. El complejo microprocesador también contiene una variedad de cofactores que incluyen la helicasa de ARN p68 (DDX5) y la p72 (DDX17), así como ribonucleoproteínas heterogéneas nucleares (hnRNPs)²⁷. Estas proteínas auxiliares, pueden tener funciones en la fidelidad, especificidad y/o en la actividad del corte que realiza DROSHA. DGCR8/PASHA funciona al menos en parte uniéndose a la región del pri-miARN que contiene la juntura entre ARNsh y ARNdH en la base de la hebilla del precursor. Esta interacción, permite que DROSHA realice el corte a una distancia de 11 pares de bases desde la base²⁸. El corte realizado por DROSHA se da de manera cotranscripcional y genera un producto con dos nucleótidos sobresalientes en el extremo 3' de la molécula. Este corte característico de este tipo de ARNasas, permite que el pre-miARN sea reconocido por Exportin 5, que transporta el sustrato hacia el citoplasma mediante un mecanismo dependiente de Ran-GTP²⁹. Una vez en el citoplasma, DICER corta los primeros 22 nucleótidos del pre-miARN con la participación de las proteínas de unión a ARNdH, llamadas *Loquacious* (LOQS) y R2D2 en *D. melanogaster*; y la proteína de unión a ARN TAR (TRBP)

junto con la proteína activadora de PKR (PACT) en mamíferos ²⁹⁻³⁵. Posteriormente, las dos hebras del miARN son separadas y una de ellas se asocia con la proteína AGO que forma parte del complejo de RISC. Una vez en el complejo de silenciamiento, el miARN sirve como guía hacia sitios complementarios dentro de ARNs blanco para mediar la represión de los mismos (Figura 3).

1.2.2 BIOGÉNESIS DE miARNs EN PLANTAS

La biogénesis de miARN en plantas difiere sustancialmente con la de animales. En plantas, ambos cortes del precursor se llevan a cabo en el núcleo, por la ARNasa III, DCL1 ³⁶. El primer corte que realiza DCL1 sobre el pri-miARN genera un precursor plegado más corto con dos nucleótidos sobresalientes en el extremo 3' denominado pre-miARN. Éste, a continuación es cortado nuevamente por la misma enzima para dar un intermediario de miARN doble hebra ³⁷ (Figura 3). Para que el proceso de biogénesis se lleve a cabo de manera correcta, DCL1 debe interactuar con la proteína de unión a ARN doble hebra, HYPONASTIC LEAVES 1 (HYL1). Otras proteínas accesorias que contribuyen con la biogénesis de miARNs son SERRATE (SE) y componentes del complejo de unión a CAP (CBC) que también son requeridos para la maduración de transcritos ³⁸⁻⁴¹. La fosfatasa CPL1, que colocaliza con las proteínas DCL1 y HYL1, participa en esta vía de una manera indirecta activando a HYL1 por desfosforilación de residuos específicos ⁴². Por otra parte, mutaciones DAWDLE ⁴³, TOUGH ⁴⁴, SICKLE ⁴⁵ y MOS2 ⁴⁶ también afectan la biogénesis de miARNs. Inmediatamente luego de la maduración, los duplex miARN-miARN* son protegidos por la acción de la enzima ARN metil transferasa HEN1, que introduce un metilo en el grupo 2'OH de la ubicado en el extremo 3' de la ribosa ⁴⁷. A continuación, esta molécula es transportada al citoplasma por el transportador HASTY (HST), que es un homólogo del transportador de miARN de animales Exportin 5 ⁴⁸. Una vez en el

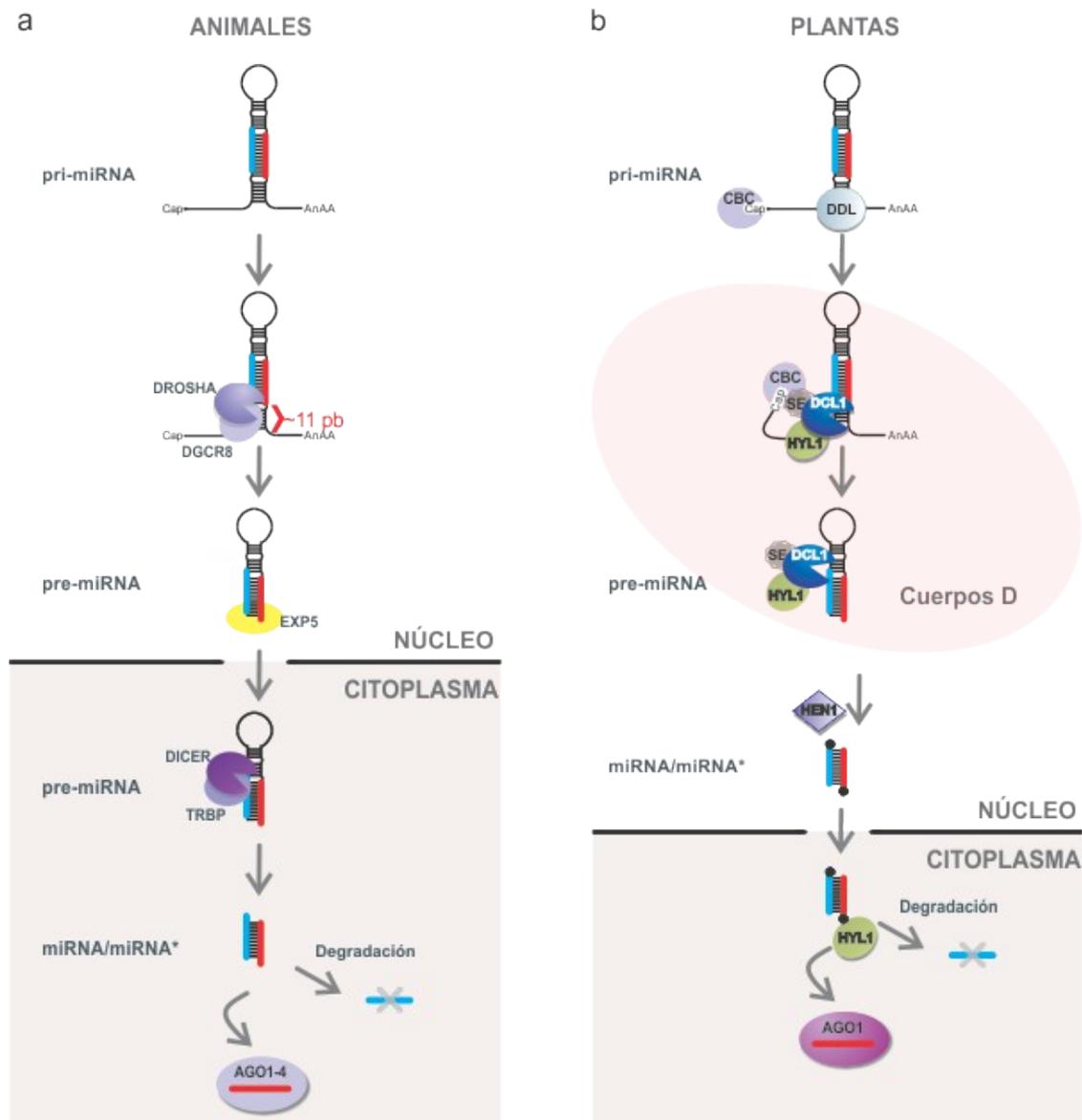


Figura 3: Biogénesis de miARNs en (a) animales y (b) plantas

citoplasma, una hebra del duplex se asocia con AGO y sirve como guía para el complejo RISC, que corta ARNs complementarios al miARN en el centro de la región apareada ²¹.

1.2.2.1 PROTEÍNAS DE PROCESAMIENTO DE PLANTAS

El reconocimiento de ARN por las proteínas encargadas de su procesamiento juega un rol esencial en la biogénesis de miARN. En plantas, la

proteína ARNasa III que cumple la función principal en la síntesis de miARN es la proteína tipo DICER, DCL1, que realiza cortes en el ARNd_h para liberar el producto final. La precisión del corte, se ve fuertemente mejorada ante la presencia de las proteínas accesorias HYL1 y SE.

1.2.2.1.1 DICER

Las proteínas DICER son endorribonucleasas multidominio de gran tamaño que juegan una función central en las vías de miARNs y de ARNi. Las proteínas DICER se encuentran presentes en la mayoría de los eucariotas, con un número de copias y estructura de dominios que varía entre los distintos organismos. En *Arabidopsis thaliana*, existen cuatro enzimas de tipo DICER, dos en *Drosophila melanogaster* y una en mamíferos, *Caenorhabditis elegans*, *Schizosaccharomyces pombe* y en otras especies de levaduras⁴⁹. Las proteínas DICER de eucariotas inferiores, generalmente tienen una organización de dominios poco compleja, con un dominio PAZ, seguido de dos dominios ARNasa III⁵⁰. En contraste, las DICER típicas de eucariotas contienen un dominio DEAD helicasa en el extremo N-terminal, seguido de un dominio de función desconocida (DUF283 o dsRBF en plantas), un dominio PAZ, dos dominios ARNasa III y al menos un dominio de unión a ARNd_h (dsRBD) (Figura 4). Las funciones de algunos de los dominios de DICER no están completamente establecidas. Los dominios PAZ y ARNasa III de DICER humana han sido extensamente estudiados y se ha determinado que están involucrados en la unión a ARN y/o en el clivaje del sustrato, actuando como una regla molecular determinando la longitud del producto final⁵⁰⁻⁵³. El dominio helicasa, participa en la selección de sustrato, debido a que es capaz de distinguir entre los duplex perfectos de ARN y aquellos que tienen estructura de hebilla, ya que tiene una mayor afinidad por estos últimos. Además, juega un rol muy importante en la catálisis, manteniendo a DICER en un estado semi-inhibido⁵⁴⁻⁵⁶. Este dominio también es mediador en la interacción entre DICER y su par, TRBP

de *Homo sapiens*. Se ha observado que esta interacción, estimula la actividad de la proteína, lo que sugiere que las proteínas accesorias podrían modular la inhibición ejercida por la helicasa^{33,35,57,58}. Por último, los roles de los dominios DUF/dsRBF y dsRBDs son menos conocidos hasta el momento^{49,59,60}.

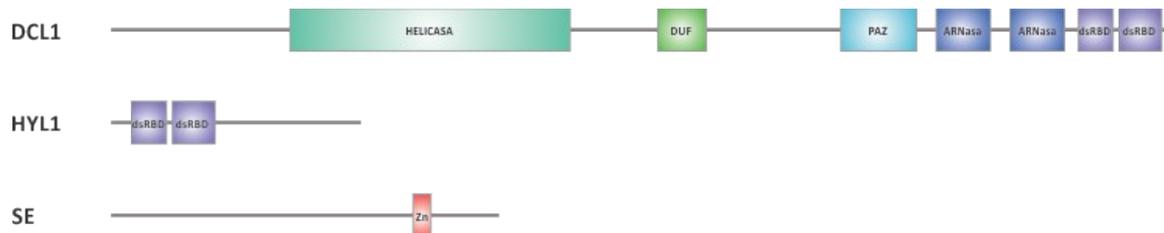


Figura 4: Distribución de dominios de las proteínas involucradas en la biogénesis de miARN en plantas. Los dominios tipo dsRBD se muestran como rectángulos de color violeta

En plantas, existen cuatro enzimas de tipo DICER llamadas DCL1-4. Cada una de ellas tiene funciones específicas en las distintas vías de generación de las moléculas efectoras del ARNi, sin embargo, existe redundancia funcional entre los miembros de esta familia. De estas cuatro enzimas, DCL1 es la principal y única responsable de la generación de miARNs. La pérdida de función de la proteína en plantas, resulta letal para el embrión de *A. thaliana*, lo que sugiere que no existe otra DCL que sea capaz de procesar eficientemente a los precursores de miARNs con estructuras de hebilla o de localizarse en el tejido adecuado al momento de cumplir su función esencial^{61,62}.

DCL1 es una proteína de 1906 aminoácidos, su estructura de dominios es similar a la de las proteínas DICER canónicas, con la diferencia que presenta un dsRBD extra en el extremo C-terminal (DCL1-B). A diferencia de las proteínas DICER de mamíferos, que ejecutan un único ciclo catalítico para liberar el miARN/miARN* de los precursores, DCL1 debe ejecutar al menos dos cortes para producir los duplex de ARN pequeños. En este sentido, DCL1 realiza simultáneamente las funciones que en animales llevan a cabo los complejos

proteicos formados por DROSHA y DICER. DCL1 corta pri-miARNs de manera procesiva, y de este modo, es capaz de generar ARN pequeños a partir de los precursores aún en ausencia de HYL y SE. Sin embargo, se ha demostrado que tanto la velocidad de reacción como la precisión en el corte se ve altamente mejorada con la presencia de las proteínas auxiliares ⁶³. A partir de estudios realizados con microscopía de fluorescencia, se observó que tanto DCL1 como HYL1 colocalizan en sitios bien definidos dentro del núcleo de las células vegetales llamados cuerpos D (*D-bodies*, en inglés), mientras que la localización de SE no está restringida a estos sitios puntuales, sino que es más heterogénea. Esta diferencia en la localización de SE respecto a las otras dos proteínas, sugiere que podría estar involucrada en otras funciones además del procesamiento de precursores de miARNs ⁶⁴, mientras que HYL1 sí sería específica de esta vía.

Los sustratos de DCL1, pre- y pri-miARN son oligonucleótidos de ARN cuya característica común es su estructura secundaria en forma de hebilla. Estos precursores no poseen una secuencia común que pueda ser reconocida ^{11,65}. A su vez, sus tamaños y estructuras son mucho más heterogéneos que los de los precursores de animales ^{66,67}. A pesar de esto, DCL1 junto con las proteínas accesorias de procesamiento, son capaces de liberar con precisión la secuencia específica del miARN a partir del pri-miARN ³⁷.

1.2.2.1.2 HYL1

HYL1 es una proteína que pertenece a la familia de proteínas de unión a ARN. Tanto HYL1 como SE son capaces de interactuar con DCL1 en los cuerpos D, donde regulan positivamente la eficiencia y precisión de procesamiento de los miARN ^{63,64,68-70}. Las plantas con mutaciones en HYL1 presentan un fenotipo pleiotrópico, con disminución en el crecimiento radicular, hojas hiponásticas (curvatura hacia arriba), retardo en el crecimiento, mayor cantidad de brotes

laterales, menor fertilidad y alteración en la respuesta a hormonas (Figura 5) ⁷¹. Las plantas con fenotipo *hyl1*, presentan acumulación de precursores de miARN y niveles bajos de miARN, lo que indica que la ausencia de una proteína HYL1 funcional afecta el sistema a nivel del procesamiento ^{38,72-74}.

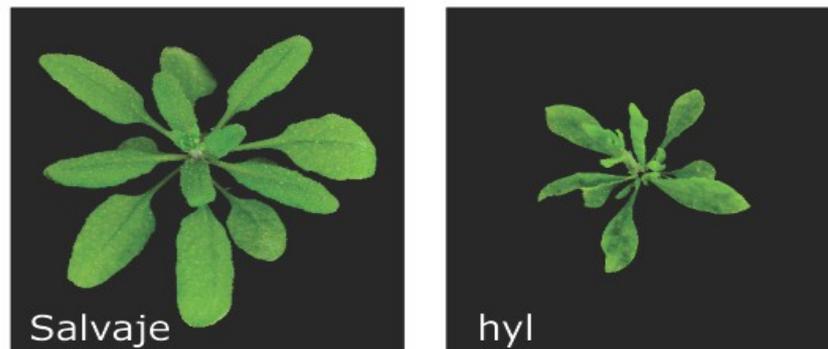


Figura 5: Comparación fenotípica de plantas de *Arabidopsis thaliana*. (A) Forma silvestre de la planta ecotipo Columbia-0. (B) *Arabidopsis thaliana* con genotipo *hyl1*.

Estructuralmente, HYL1 contiene dos dominios de unión a ARN doble hebra en el extremo N-terminal (residuos 1-170), seguido por una señal de localización nuclear (NLS) y una región desestructurada hacia el extremo C-terminal que contiene una secuencia de seis repeticiones de 28 aminoácidos (residuos 171-419). Se ha demostrado tanto *in vivo* como *in vitro* que los dos dominios dsRBD de HYL1 son necesarios y suficientes para el funcionamiento correcto de la proteína ⁷⁵ y que el primero sería el principal responsable del reconocimiento de sustrato ^{76,77}. Estudios bioquímicos y estructurales revelaron que el primer dsRBD de HYL1 confiere la capacidad para unirse al ARNdh, mientras que el segundo participaría en interacciones proteína-proteína. Si bien HYL1 es una proteína específica de plantas, se ha sugerido que podría participar en la selección de hebra del duplex de miARN y la carga de la misma en el complejo RISC de manera análoga a TRBP de animales ⁷⁸, con la que comparte

parcialmente la arquitectura de dominios.

1.2.2.1.3 SERRATE

SE es un componente esencial para el procesamiento eficiente y preciso de miARNs en plantas. Los alelos nulos de *se* son letales para el embrión, mientras que alelos mas débiles son pleiotrópicos y exhiben hojas aserradas, hiponásticas y plantas con fertilidad reducida. En *Arabidopsis*, SE controla el desarrollo de hojas, la actividad meristemática y la arquitectura de las inflorescencias^{74,79,80}. Al igual de lo que ocurre con plantas deficientes en HYL1, los niveles de miARN en plantas mutadas en el gen *se* son bajos respecto a los de plantas silvestres, mientras que los niveles de pri-miARNs están aumentados^{68,70}.

La arquitectura de dominios de la proteína SE consiste en múltiples señales de localización nuclear en el extremo N-terminal, un dominio Mid, un dominio dedos de zinc de tipo C2H2 y un tallo desestructurado hacia el extremo C-terminal. En animales, plantas y hongos, existen genes homólogos a *se*, sin embargo, la similitud entre ellos no es demasiado grande⁸⁰. Una mutación en el gen *se* de pez cebra tiene un fenotipo letal para el embrión, indicando que SE tiene una función crucial durante el desarrollo en animales⁸¹.

SE es capaz de unir ARN doble hebra a través del dominio dedos de zinc y el extremo C-terminal desestructurado, pero también cumple una función guiando al sustrato hacia la maquinaria de procesamiento, ya que interacciona con la proteína del complejo macromolecular, HYL1⁸². SE se expresa en el núcleo, al igual que DCL1 y HYL1, sin embargo, a diferencia de estas proteínas, su expresión no es constitutiva, sino que está limitada a tejidos y a etapas del desarrollo específicas⁸⁰. Posiblemente SE funcione como una plataforma de reclutamiento de sustrato y de proteínas de procesamiento, y su expresión diferencial sea responsable de algún tipo de regulación de los miARN.

1.2.2.1.4 CBC

El complejo nuclear de unión a cap (CBC), es un heterodímero formado por dos subunidades, la proteína de unión a cap 80 (CBP80) y CBP20⁸³. En *Arabidopsis*, CPB80/ABA HYPERSENSITIVE1 (ABH1) fueron identificadas a partir de una búsqueda de mutantes con respuesta alterada a ácido abscísico (ABA)⁸⁴. *Cpb80/abh1* confiere hipersensibilidad al ABA e incrementa la resistencia a sequía⁸⁴. Los alelos nulos de *cbp80/abh1* y *cbp20* presentan un fenotipo con hojas aserradas, que es similar a los alelos débiles de *se*^{74,80,84,85}, indicando que el complejo CBC y SE podrían tener funciones solapadas. En efecto, *cbp80/abh1* y *cbp20* causan defectos en los cortes de los pre-miARNs, reducen los niveles de miARNs y aumentan los niveles de pri-miARNs, demostrando que al igual que SE, el complejo CBC está involucrado tanto en los cortes del pre-miARN como de la biogénesis de miARNs^{41,86}. Estos resultados, también sugieren una función de la estructura del cap, facilitando el procesamiento del pri-miARN además de su función relacionada al control de la estabilidad del precursor^{26,87}.

1.3 REGULACIÓN DE LA VÍA DE miARNs

Por tratarse de moléculas reguladoras esenciales, los niveles de miARNs deben ser estrictamente regulados para mantener la homeostasis celular. La desregulación del proceso de biogénesis de miARNs, altera la cantidad de moléculas efectoras, lo que modifica los niveles de sus blancos y en última instancia, afecta severamente el metabolismo celular. A pesar de su relevancia en el desarrollo celular, hasta el momento no se ha demostrado la existencia de algún tipo de regulación específico sobre el procesamiento de precursores de miARNs de plantas. Sin embargo, algunos estudios, han aportado información valiosa sobre algunas formas de regulación transcripcional y post-transcripcional de las proteínas de procesamiento. Por un lado, se han encontrado en los

promotores de muchos genes de miARNs de plantas, motivos de unión tejido específicos y para factores de transcripción regulados por estrés ⁸⁸. Por otro lado, las proteínas DCL1, AGO-1 y AGO-2 de *A. thaliana* son reguladas por un mecanismo de retroalimentación negativa. Los genes que codifican para estas proteínas, poseen sitios blancos de miARNs. Por ejemplo, el miR162 y el miR414 inhiben la expresión del gen *dcl1* ^{89,90}. La presencia de estos sitios blanco, provoca que cambios temporales o espaciales en la expresión de estos miARNs, efecten los niveles de DCL1 y por lo tanto los niveles globales de los miARNs. Los mismos principios se pueden aplicar en la regulación de AGO1 ^{91,92} y AGO2 ¹⁶, que son regulados por el miR168 y el miR403, respectivamente. En particular la homeostasis de AGO1 es alcanzada por varios ciclos regulatorios que permiten que las rutas de los miARN y siARN funcionen correctamente. Estos ciclos incluyen el corte de ARNm de AGO1 mediado por miR168 ⁹², el aumento en la acumulación de los niveles de miR168 cuando se incrementan los niveles de AGO1 ⁹ y la represión traduccional de AGO1 mediada por AGO10 ⁹³.

Se ha demostrado que la velocidad de degradación de los miARNs es una forma importante de regulación de su actividad. Por ejemplo, en *A. thaliana* el "apagamiento" de genes que codifican para nucleasas SDN (del inglés, *small RNA dependent nucleases*), encargadas de cortar miARNs, llevan a la acumulación de estas moléculas pequeñas y a la presencia de numerosos defectos en el desarrollo de las plantas ⁹⁴. Nucleasas similares en nemátodos cortan ARN pequeños, incluyendo miARNs, en la dirección 5' a 3', mientras que en plantas las nucleasas SDN actúan en la dirección opuesta ^{95,96}. La presencia de la cola poli(U) en el extremos del miARN, inhibe su degradación por parte de las nucleasas SDN ⁹⁷. Otro ejemplo de regulación del procesamiento general fue descrito en un trabajo en el que se demostró que la sobreexpresión en *Arabidopsis* de un elemento de ARN denominado SINE (del inglés, *short interspaced elements*)

genera fenotipos que se asemejan a mutantes en el procesamiento. Se demostró que las estructuras de tallo y burbuja de SINE imitan estructuras de precursores de miARNs y se unen a HYL1⁹⁸. Estos resultados sugieren que la expresión específica del ARN SINE en un determinado tejido o estadio del desarrollo podría modular la producción de miARNs por competencia con HYL1.

1.4 RECONOCIMIENTO DE SUSTRATO EN LA BIOGÉNESIS DE miARNs

1.4.1 DETERMINANTES ESTRUCTURALES DE LOS PRECURSORES DE miARN

Una característica distintiva de las enzimas tipo DICER es que son capaces de digerir de manera procesiva moléculas de ARN doble hebra, liberando fragmentos de 22 pb que corresponden al par miARN/miARN* o siARN/siARN*⁵¹. Por esta razón, la secuencia del miARN maduro va a estar definida por la posición en que se realice el primer corte en el precursor, dado que el segundo corte es realizado por DICER a una distancia de 22 pb del extremo. En el procesamiento de miARN, el posicionamiento adecuado del primer corte define el registro de procesamiento que determina la secuencia del producto final. El miARN es un mensaje de solo 22 letras, por lo que un desplazamiento en la digestión del precursor altera sustancialmente su secuencia y por consiguiente el efecto que tiene el producto en la regulación post-transcripcional.

La estructura de los precursores de miARN juega un papel muy importante en el mecanismo de procesamiento⁶⁶. Los precursores de animales se caracterizan por presentar una estructura con un tallo imperfecto de una longitud de ~65 nt, seguido por segmentos inestables de ARN en sus extremos. Estos segmentos son: un bucle terminal hacia un extremo y ARNsh hacia el otro. De esta manera, el miARN animal puede ser dividido en cinco partes: un bucle terminal, un dúplex, miARN/miARN*, un tallo inferior y secuencias flanqueantes simple hebra (segmentos basales). El miARN/miARN* posee una longitud de ~22

pb (dos vueltas de hélice). En animales, la mayor parte de la especificidad en la selección de la secuencia del miARN dentro del precursor está dada por el primer corte realizado por DROSHA-DGCR8, que actúa como una regla molecular, cortando a aproximadamente 11 nucleótidos desde el comienzo de la estructura de hebilla ^{28,99}. Esta distancia es constante en los precursores caracterizados hasta el momento, y determina la posición del miARN en el pri-miARN y, por lo tanto, su secuencia. Los precursores de miARNs de plantas constituyen un amplio rango de estructuras, son muy heterogéneos y sus tamaños varían entre 50 a 900 nt ^{66,100}. Sin embargo, ha sido posible identificar distintas familias de precursores que comparten características estructurales comunes y que, en última instancia, se asocian a distintos mecanismos de procesamiento. El primer mecanismo de procesamiento se denominó **mecanismo corto desde la base al bucle**. Ocurre en precursores que contienen una región con una protuberancia, seguida de un tallo menor de aproximadamente 15-16 bases parcialmente desapareadas ubicado debajo de la secuencia del miARN/miARN*. En esta región de reconocimiento, las bases consecutivas desapareadas nunca son más de 3 ¹⁰¹⁻¹⁰⁴. En un trabajo realizado recientemente, se demostró que existe otro tipo de procesamiento similar que involucra a la familia de MIR169. Los precursores de estos miARNs, tienen un sitio de corte ubicado a 21 nt desde la base proximal del miARN/miARN*, por lo que necesitan un corte adicional para que se libere la molécula efectora. Este tipo de procesamiento se denominó **mecanismo largo desde la base a bucle** ¹⁰⁴. La diferencia principal con el mecanismo de base a bucle canónico radica en que estos precursores necesitan tres eventos de corte donde el miARN es liberado por el segundo y tercer corte. Durante este tipo de procesamiento, además de producir el miARN, se generan ARN pequeños adicionales que generalmente se acumulan en bajos niveles. En ambos casos el determinante estructural reconocido es un tallo menor con bases parcialmente

apareadas. Sin embargo, esta estructura no está presente en todos los precursores de plantas ¹⁰¹. Otra familia de precursores que incluye a los MIR319 y MIR159 no presentan ninguna región estructurada en la parte inferior del miARN, pero se caracterizan por tener un tallo superior conservado extenso que se procesa **desde el bucle a la base**. En este mecanismo, DCL1 realiza cuatro cortes secuenciales ^{66,105} generando otros ARN pequeños ^{106,107}. Recientemente, se ha identificado un grupo de precursores que tiene una región terminal de longitud conservada (~42 nt) y un pequeño bucle que. Este grupo de precursores se procesa por un **mecanismo corto desde el bucle a la base**, donde la maquinaria de procesamiento es guiada por un segmento del tallo superior y a partir de dos cortes, se libera el miARN ¹⁰⁴.

1.5 DOMINIOS dsRBDs

Durante el procesamiento de miARNs participan distintas proteínas. Una característica distintiva de las mismas es la capacidad de unir ARNdH. En este sentido, los dominios de unión a ARN doble hebra cobran un rol sumamente importante tanto para el reclutamiento del sustrato como para el posicionamiento correcto de los sitios activos. En conjunto las proteínas que forman el complejo de procesamiento de plantas poseen siete dominios con capacidad de unir ARN ¹⁰⁸. Cuatro de ellos pertenecen a la familia de dominios dsRBD, de los cuales se ha demostrado que reconocen específicamente variantes de motivos de ARN doble hebra ¹⁰⁹. Un quinto dominio, el dominio DICER-dsRBF, posee un plegamiento de tipo dsRBD, aunque aparentemente no une ARN y estaría involucrado en la unión a otras proteínas ¹¹⁰. Los restantes son un dominio PAZ en DCL1, cuyos homólogos han sido extensamente caracterizados en diferentes proteínas, y un dominio tipo dedo de Zn(II) en SERRATE (Figura 4).

Los dominios dsRBD tienen la particularidad de reconocer variantes de

segmentos de ARN doble hebra ¹⁰⁹. Estos dominios adoptan un plegamiento común $\alpha\beta\beta\alpha$, donde las dos hélices α se apoyan sobre una misma cara de la lámina β . La interacción con ARN doble hebra ha sido caracterizada a partir de la resolución de estructuras de complejos dsRBD-ARNdh ^{111–115}. En dichas estructuras, se pudo observar que durante la interacción, la proteína se apoya sobre una cara de la hélice del ARNdh y se extiende 16 pb del sustrato, lo que constituye una vuelta y media de hélice. El patrón de interacción de los dsRBDs está definido por tres regiones. La **región 1**, ubicada en la primera hélice, donde hace contacto con el surco menor del ARN; la **región 3** definida por el bucle 4 y parte de la hélice 2, que hace contactos con el surco mayor; y la **región 2**, formada por el bucle 2 que interacciona con el surco menor de la misma cara de la hélice que la región 1 (Figura 6).

Si bien está establecido que la principal función de los dsRBDs es el reclutamiento de moléculas de ARN hacia una proteína multidominio, se ha observado que los mismos, son capaces de interactuar con distintas clases de macromoléculas. Por ejemplo, los dsRBDs de STAUFEN tienen funciones de unión a proteína y ARN y las mismas están claramente separadas en los distintos dsRBDs. Los dsRBD 1, 3 y 4 muestran capacidad de interactuar con el ARN mientras que dsRBD 5 y dsRBD 2 actúan como motivos

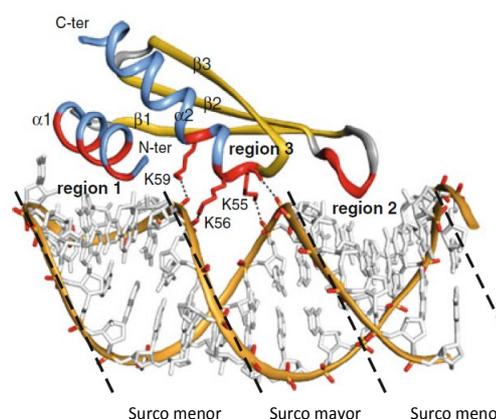


Figura 6: Vista global de la superficie de unión al ARN canónica del complejo ARN y el dsRBD1 de ADAR2 (pdb 2L3C). La región 1 (hélice $\alpha 1$) y la región 2 (bucle 2) se insertan en dos surcos menores sucesivos. La región 3 (bucle 4) contacta el esqueleto fosfodiéster del surco mayor del ARN ²¹².

de interacción con proteínas. Sin embargo, la unión al ARN y proteína no son mutuamente excluyentes. Además de la unión al ARNdh, los dsRBDs de PKR aislados son capaces de formar heterodímeros con PKR completa a partir de la interacción entre sus dsRBDs ¹¹⁶. Otras interacciones proteína-proteína han sido observadas en proteínas PACT y TRBP, cuyos dsRBDs son capaces de formar heterodímeros independientes de ARN con los dsRBDs de PKR y de esta manera modulan su actividad ^{117,118}. Los dsRBDs no sólo son capaces de interactuar con otros dsRBDs, sino que lo hacen con dominios de otro tipo en forma inter o intramolecular. Se ha demostrado que el dsRBD2 de PKR es capaz de interactuar con el dominio quinasa de la misma proteína. Esta interacción bloquea la accesibilidad de la proteína hacia el sustrato y de esta manera se produce su autoinhibición ¹¹⁹. Una interacción similar se ha observado entre el dsRBD y el dominio catalítico de la ARNasa III de *Aquifex aeolicus* ¹¹³. También existen interacciones entre dsRBDs y dominios no catalíticos. El dsRBD5 de STAUFEN de *D. melanogaster* interactúa con la proteína MIRANDA, que media la localización de proteínas y ARN en el sistema nervioso durante el desarrollo ¹²⁰. A pesar de ser una característica común a varios dsRBD, hasta el momento no se ha caracterizado estructuralmente ningún complejo dsRBD:proteína, ni se conoce si existen regiones o residuos que participen preferencialmente en este tipo de interacciones. Además de unir ARNdh y proteína, algunos de estos dominios tienen la capacidad de unir también moléculas de ADN. Los primeros 250 aminoácidos de RHA contienen dos dsRBDs y en esa misma región se ha reportado una actividad de unión a ADNdh ¹²¹. Hasta el momento, no está muy claro la forma en que RHA puede unir tanto la hélice A de ARN como la hélice B de ADN, aunque un estudio de mutagénesis, sugiere que la unión puede involucrar dos grupos de residuos distintos pero solapados, que incluyen a el primer dsRBD y una extensión del extremo C-terminal. Otro dsRBD capaz de unir

tanto ARNdh como ADNdh es el dominio no canónico de *S. pombe*. Si bien hasta el momento no está establecido el mecanismo de reconocimiento se sospecha que las características estructurales particulares de este dominio (extensión C-terminal, localización inusual de la hélice $\alpha 1$ e inserción en el bucle $\beta 1$ - $\beta 2$) serían los responsables de esta capacidad de unión dual.

Además de interactuar con ácidos nucleicos y proteínas, se ha demostrado que este pequeño dominio es capaz de cumplir con otras funciones. XLRBPA y su dsRBD aislado, así como el primer dsRBD de PKR, poseen actividad de anillado de hebras de ARN ^{122,123}. Esto es consistente con el descubrimiento de que los dsRBDs de PKR son capaces de estirar el ARN que contiene bases desapareadas o protuberancias ¹²⁴, y también puede ayudar a explicar algunos efectos de cooperatividad en la unión de ARNdh por parte de los dsRBDs ubicados en tandem. Quizás la actividad tipo chaperona de estos dsRBDs que son deficientes en la unión al ARN y proteínas puede facilitar el anillado de las hebras y el repliegamiento de sustratos de ARN para un correcto reconocimiento por dsRBDs vecinos.

Otras funciones de los dsRBDs se relacionan con la regulación de la actividad catalítica de la proteína que los contiene. PKR es un componente vital para el mecanismo celular antiviral y la unión a ARNdh, que se sintetiza en grandes cantidades durante una infección viral, puede resultar en la dimerización de la proteína y la subsecuente autofosforilación y activación de la misma ¹²⁵. La actividad quinasa de PKR es inhibida por un ARN codificado por el virus y por inhibidores proteicos tales como el ARN VA del adenovirus y la proteína portadora de un dsRBD, E3L del virus de la vacuna ^{119,126}. Dicha función es también fuertemente regulada por factores celulares tales como el activador proteico, PACT y el inhibidor, TRBP que contienen tres dsRBDs cada uno ^{117,118}. Esta red de interacciones está basada en algunos contactos entre dsRBDs;

dsRBDs-quinasa y dsRBD-ARNdh, y su núcleo está formado por una unidad regulatoria formada por los dsRBDs de PKR en tandem. De manera similar, estas interacciones complejas ocurren en los tres dsRBDs de la enzima de edición de ARN (ADAR1) ¹²⁷.

Otra función que desempeñan los dsRBDs está relacionada a la localización, ya que son capaces de actuar como señales de localización nuclear atípicas. Un ejemplo de esto es el tercer dsRBD de ADAR1, y el dsRBD no canónico de DICER de *S. pombe* ¹²⁸⁻¹³⁰. Recientemente se ha reportado que el dsRBD de DICER humana actúa como una NLS no clásica cuando se expresa el dsRBD fusionado a un gen reportero, sin embargo, esta función es inhibida cuando el dominio se encuentra en el contexto de la proteína completa. En este mismo trabajo, se demostró también que no todos los dsRBDs son capaces de actuar como NLSs, lo que indica que el direccionamiento nuclear no es una propiedad general de este tipo de dominios ¹³¹.

1.5.1 CARACTERÍSTICAS ESTRUCTURALES DE LOS dsRBDs

El dsRBD es un dominio de ~65-70 aminoácidos que está presente en eucariotas, procariotas y en proteínas virales. Este dominio, que interacciona principalmente con el ARNdh fue inicialmente reconocido en 1992 como un dominio funcional conservado. Esto se logró por homología de secuencia entre tres proteínas diferentes: STAUFEN, una proteína responsable de la localización del ARNm en *Drosophila*; TRBP, una proteína multifuncional que fue inicialmente caracterizada como un activador de la expresión del gen HIV-1; y la proteína A de *Xenopus laevis* (XLRBPA), que es un homólogo de TRBP ¹³². En el mismo estudio, la búsqueda en bases de datos permitió la identificación de otras proteínas que contenían una copia del dominio. Al mismo tiempo, diferentes dominios fueron caracterizados como dominios de unión a ARNdh en proteínas como PKR, una

proteín quinasa dependiente de ARNdh¹³³⁻¹³⁵. En la figura 7 se muestra un alineamiento de secuencias de dsRBDs de distintos organismos con la secuencia consenso debajo. Como puede observarse, la región C-terminal es la más conservada. Los dos primeros tercios más divergentes, y en algunos casos, donde la variación en esta región es demasiado elevada, los dsRBD se clasifican como dsRBD tipo B. Se ha demostrado que los dsRBD de este tipo, presentan una menor actividad para unir el ARNdh que los dsRBD con secuencias más conservadas en el extremo N-terminal^{132,136}.

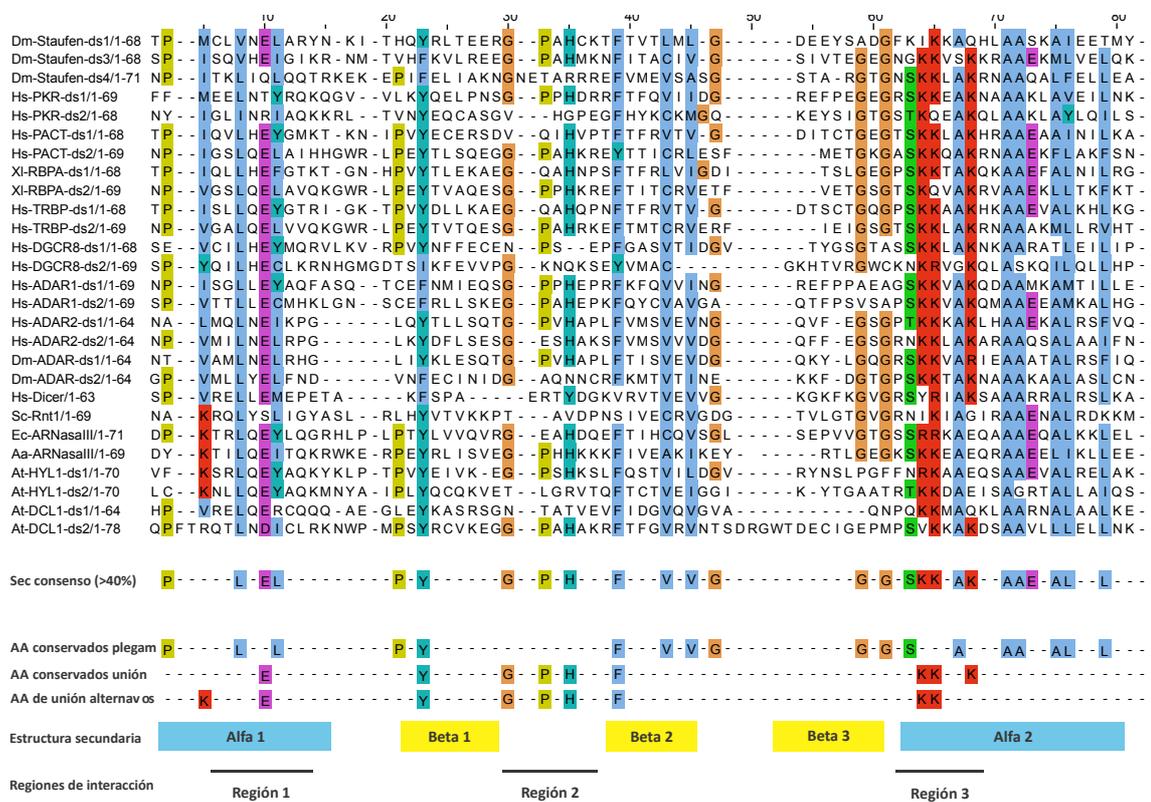


Figura 7: Alineamiento de secuencias de dsRBDs de *Homo sapiens* (Hs), *D. melanogaster* (Ds), *S. cerevisiae* (Sc), *X. laevis* (Xl), *A. thaliana* (Ath) y *E. coli* (Ec). El alineamiento está coloreado por conservación (>40%) y propiedades de los aminoácidos (AA). Debajo del alineamiento se muestran los AA conservados para el plegamiento y/o para la unión a ARNdh y los elementos de estructura secundaria. También se indican las tres regiones canónicas de interacción.

Las primeras estructuras tridimensionales de dsRBDs, fueron resueltas por RMN en solución, y correspondían al dsRBD de la ARNasa III de *E. coli* y al tercer dsRBD de STAUFEN de *Drosophila*. A partir de las mismas, se pudo distinguir un plegamiento mixto α/β con una topología conservada $\alpha\beta\beta\beta\alpha$ donde las dos hélices α están empacadas contra las tres láminas β antiparalelas^{137,138}. Hasta el momento, se han determinado alrededor de 30 estructuras de dsRBDs ya sea por RMN como por cristalografía de rayos X. Estas estructuras confirmaron dicha topología y las propiedades características del plegamiento de dsRBDs, aunque también, se han descubierto algunas variaciones y algunas extensiones respecto al plegamiento canónico de los mismos.

La estructura del segundo dsRBD de XLRBPA constituye el arquetipo de un dominio dsRBD canónico¹¹¹. Este dsRBD está representado en la figura 8. Los residuos conservados que coinciden con la secuencia consenso (>40 %) del alineamiento de la figura 7 están resaltados y se muestran como palos en la estructura. Estos residuos parecen ser conservados y la posible explicación a esto

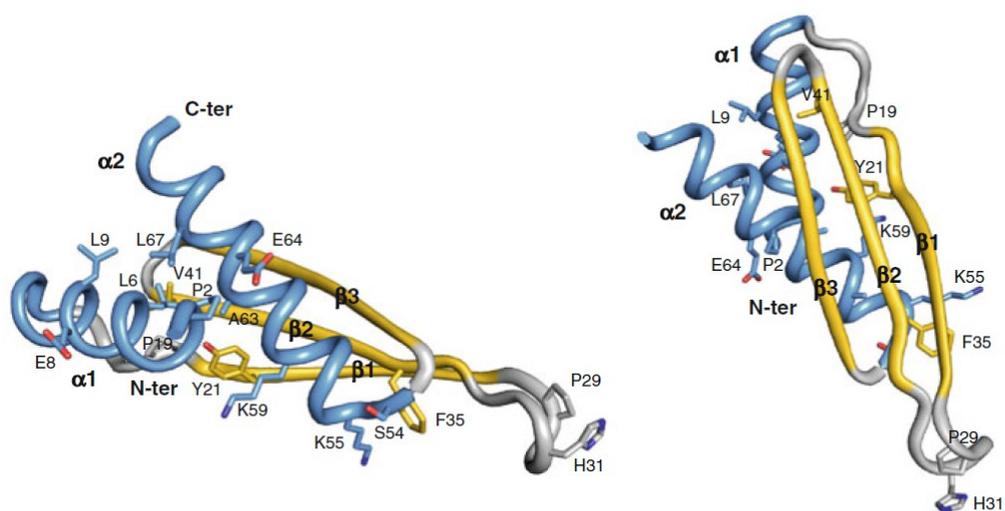


Figura 8: Estructura tridimensional del dsRBD canónico XLRBPA-ds" (pdb 1D12). En azul están representadas las hélices α y en amarillo las láminas β . Los residuos conservados están indicados como palos²¹².

sería que por un lado, mantienen estable el núcleo hidrofóbico con las dos hélices α empaquetadas sobre las láminas β y por otro lado, mantienen las capacidades de unión en forma óptima. Mediante análisis mutacionales llevados a cabo en diferentes dsRBDs, se ha demostrado que existen tres regiones que resultan importantes para la unión a ARNdh como se mencionó previamente en el apartado 1.5^{135,138–142}. En cada una de estas regiones, existen residuos o motivos conservados: la cadena lateral de E8 en la hélice α_1 , el motivo conservado GPxH en el bucle β_1 - β_2 y los residuos cargados positivamente en el motivo KKxAK al comienzo de la hélice α_2 . En estas tres regiones, la conservación de residuos puede explicarse debido a su importancia en la unión al sustrato (Figura 9).

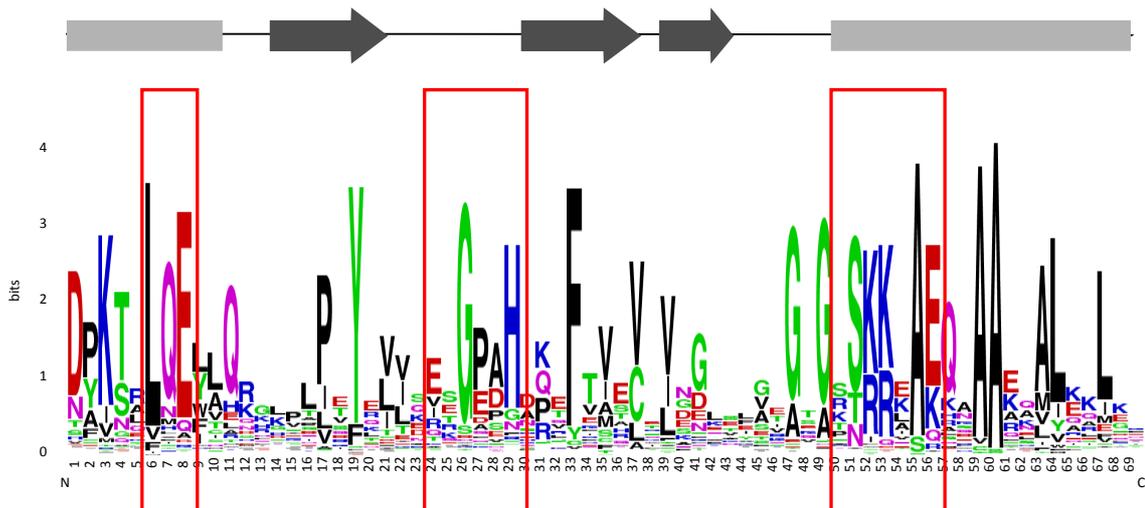


Figura 9: Logo de secuencia de prosite (<http://prosite.expasy.org/PDOC50137>). El tamaño de la letra que representa al residuo indica la probabilidad de encontrar al mismo en una posición dada dentro de la secuencia primaria en la familia. En rojo se resaltan las regiones de unión al ARNdh.

1.5.2 VARIACIONES RESPECTO A LOS dsRBDs CANÓNICOS

Aunque las estructuras de los dsRBDs están generalmente conservadas, se han encontrado algunas variaciones, especialmente en dos regiones del dominio.

La primer región de variabilidad consiste en el bucle entre la hélice $\alpha 1$ y la lámina $\beta 1$. Además, la longitud de la hélice $\alpha 1$ a veces puede ser menor como en el caso de la ARNasa III de levaduras, la ADAR2 de mamíferos y ADAR de *Drosophila*. Hasta el momento no se han encontrado propiedades particulares en términos de unión al sustrato en estos dsRBDs con hélice $\alpha 1$ más cortas^{115,143}. La segunda región variable respecto a los dsRBDs canónicos incluye el bucle que conecta a la lámina $\beta 1$ con la $\beta 2$. Este bucle que generalmente tiene seis aminoácidos de longitud puede incorporar una inserción como es el caso de DCR1 de *S. pombe* y de HEN1 de *A. thaliana*^{144,145}. En el caso de DCR1 de *S. pombe*, este bucle más extenso es heteogéneo en la forma libre, y se ha demostrado que una mutante por delección es incapaz de unir ARNd_h¹⁴⁴. En la estructura cristalográfica de HEN1 de *A. thaliana*, se ha encontrado que los residuos conservados en este segundo bucle más extenso, interaccionan con otros residuos conservados del dominio catalítico metil transferasa¹⁴⁵. Esto sugiere que los bucles $\beta 1$ - $\beta 2$ más extensos podrían tener un papel dual en la unión a ARNd_h y en unión a proteínas orientando los dominios catalíticos con respecto al sustrato. Además de todas estas variaciones, también se han observado algunas extensiones notables. Hasta el momento, todas se dan en el extremo C-terminal. Tal es el caso de la ARNasa RNT1P de levaduras que presenta una tercera hélice en esta posición. Se ha propuesto que esta extensión participaría indirectamente en la unión al ARN ayudando a posicionar la hélice 1, principal determinante de la unión, en el sitio correcto^{114,143}. Una variación similar ocurre con DCR1 de *K. polysporus*, que presenta una tercer hélice más corta que la de RNT1P. Hasta el momento, no se le ha adjudicado ninguna función. El caso más notable de extensiones C-terminales de dsRBDs es el de DCR1 de *Schizosaccharomyces pombe*. Esta proteína presenta un pequeño giro de hélice α seguido por un sitio de coordinación a zinc. Se ha determinado que esta extensión C-terminal es crítica

para la localización nuclear de DCR1^{130,144}. Como se mencionó en el apartado 1.5, existen otros reportes donde se ha propuesto que algunos dsRBDs median el tráfico nucleocitoplasmático, promoviendo tanto, la importación como la exportación al núcleo^{129,146-149}, sin embargo, los mecanismos que controlan la difusión núcleo-citoplasmática son poco conocidos.

1.5.3 CARACTERÍSTICAS ESTRUCTURALES DEL SUSTRATO DE dsRBDs

La capacidad más sobresaliente de los dsRBDs es su capacidad para discriminar entre la variedad química y estructural de polímeros de ácidos nucleicos para unir preferencialmente ARNdH. Los primeros estudios realizados en sustratos de la ARNasa III de levadura, mostraron que la secuencia del ARNdH no era un factor crucial para el reconocimiento. Sin embargo, algunas variaciones en la secuencia del mismo, podían afectar la afinidad de la unión. Los dsRBDs presentan una preferencia absoluta por sustratos con conformación de hélice A. Estudios bioquímicos han demostrado que la longitud de la hélice requerida para que se lleve a cabo la unión es heterogénea y oscila entre los 12 a 16 pares de bases. El efecto de interrumpir esta hélice, varía enormemente dependiendo de la naturaleza de dicha modificación^{139,150}. En algunos casos se puede llegar a impedir la unión completamente¹³⁹, mientras que en otros, sólo se produce una leve disminución de la afinidad¹⁵¹.

La conformación que el ARNdH adopta en solución es la conocida como forma A de la doble hélice y se diferencia de la hélice B del ADNdh en varios aspectos. Por un lado, la hélice A del ARN tiene un diámetro de 23 nm, mientras que la de ADN tiene aproximadamente 20 nm. Una vuelta de hélice de ARNdH se completa con 11 pb, mientras que una de ADN sólo necesita 10 pb (Figura 10).

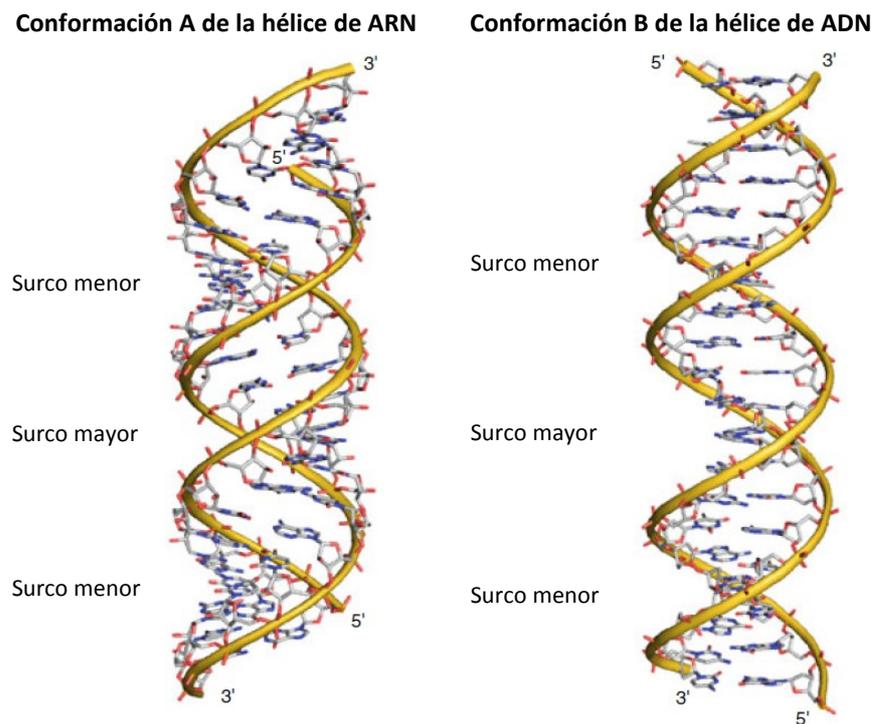


Figura 10: Diferencias en la topología de la hélice A del ARNdh y la hélice B del ADNdh. Notar especialmente las diferencias entre la dimensión y accesibilidad del surco menor y surco mayor de ambas hélices ²¹².

Por otro lado, en la conformación A que adopta el ARNdh en solución, los planos de los pares de bases están inclinados con respecto al eje de la hélice, de modo que no se ubican de manera perpendicular. El surco mayor es angosto y profundo, lo que obstaculiza el acceso a las bases, mientras que el surco menor es ancho y superficial. Debido a la diferencia en la accesibilidad de los surcos, los contactos entre las bases del ARN y los dsRBDs ocurren principalmente en el surco menor. Sin embargo, en esta región el reconocimiento de secuencias específicas no es posible debido a la conformación de las bases en la hélice A del ARN. Por otra parte, la conformación de los azúcares en las hélices A de ácidos nucleicos, es C3'-endo, mientras que en las conformaciones de hélice B es C2'-endo. La razón por la que las hélices de ARNdh se encuentran en conformación A,

se debe a que existe un impedimento estérico por la presencia de los grupos 2´OH, que no les permite adoptar la conformación B. Es decir que el ARNdh, presenta características químicas (presencia o ausencia de 2´-OH) y características estructurales (ancho del surco mayor y menor) que permiten discriminar al ARNdh del ADNdh.

Un punto clave en la comprensión molecular del reconocimiento por parte de los dsRBDs es la manera en que se obtiene la especificidad estructural para la doble hélice de ARN que corresponde a un blanco celular específico. En algunos casos, la especificidad puede ser lograda a partir de la interacción coordinada de múltiples copias de dsRBDs en tandem dentro de una proteína multidominio. Dicha interacción, puede explicar la unión cooperativa al ARNdh por PKR y otras proteínas multidominio. Sin embargo, estudios realizados en una gran cantidad de proteínas que contienen dsRBDs ^{127,152-154} mostraron que un sólo dsRBD es suficiente para brindarle a la proteína la especificidad para la selección del sustrato blanco. Por el contrario el reconocimiento de secuencias o elementos de estructura secundaria que flanquean la hélice podrían ser las claves de la especificidad.

1.5.4 DESCRIPCIÓN DE LA INTEFASE DE UNIÓN dsRBD-ARN

A partir de la obtención de estructuras de alta resolución de dsRBDs formando complejo con ARNdh, se ha podido obtener una descripción precisa de la superficie de unión del ARN y se ha revelado la forma en que las diferentes regiones del dominio reconocen específicamente la hélice A que adopta el ARNdh.

1.5.4.1 RECONOCIMIENTO DEL SURCO MAYOR POR LA REGIÓN 3

La secuencia de aminoácidos que corresponde a la región 3 de un dsRBD, consiste en una secuencia de aminoácidos bien conservados del motivo KKxAK,

que forma parte de la secuencia consenso de estos dominios ¹³². En un dsRBD canónico, las cadenas laterales de los aminoácidos ubicados en esta región contactan el esqueleto fosfodiéster de ambas hebras de ARN a través del surco mayor de la doble hélice. Las cadenas laterales de la primera y tercer lisina del motivo apuntan hacia una de las hebras de ARN, mientras que la segunda lisina apunta hacia la otra hebra, formando un arco que se extiende sobre el ancho del surco mayor. A nivel atómico, el grupo amino de las lisinas hace contacto con los átomos de oxígeno del esqueleto del ARN. Este conjunto de interacciones resulta de una orientación particular de las cadenas laterales de las tres lisinas que están estabilizadas por (i) la rigidez del esqueleto peptídico, que está incrustado en la estructura secundaria de α -hélice y (ii) un conjunto importante de interacciones de van der Waals entre las cadenas laterales de la primera y la tercera lisina (K55 y K59) con otras cadenas laterales del núcleo hidrofóbico del dominio. A pesar de ser el responsable de todas estas interacciones, este motivo no es absolutamente conservado. Las variaciones más frecuentes son sustituciones de una lisina por un residuo de arginina o glutamina donde se mantiene la densidad de carga positiva. Sin embargo, existen dominios donde la tercer lisina es reemplazada por un residuo cargado negativamente o una glicina. En estos casos, la pérdida de la densidad de carga positiva se compensa con una lisina en la hélice α_1 (Figura 11).

1.5.4.2 RECONOCIMIENTO DEL SURCO MENOR DEL ARNdh

El reconocimiento del surco menor del ARNdh por parte de los dsRBDs se lleva a cabo por la región 1 (hélice α_1) y la región 2 (el bucle 2 que une la primera y segunda lámina β del dominio).

La hélice α_1 , es una hélice anfipática, con una región hidrofóbica que está involucrada en la formación del núcleo hidrofóbico del dsRBD, y una región hidrofílica expuesta al solvente que hace contacto con las bases y las ribosas en

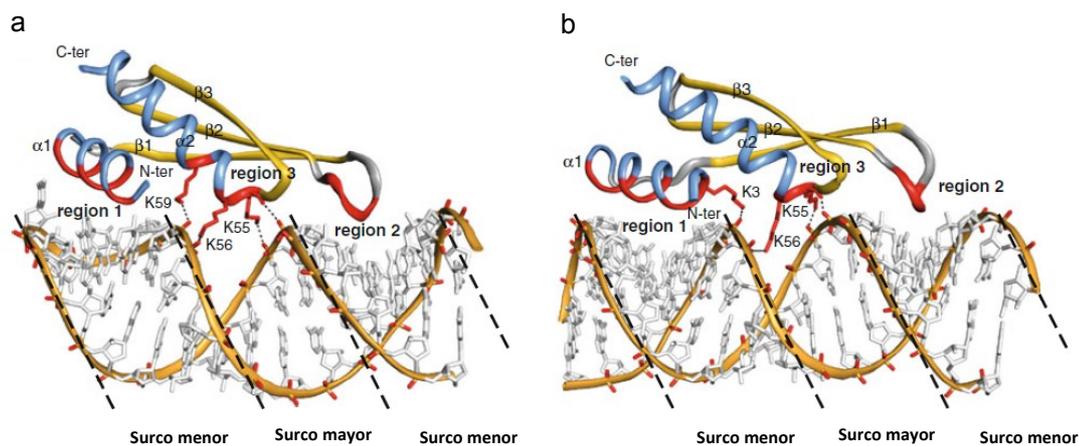


Figura 11: a) Vista global de la superficie de reconocimiento de ARN canónica del complejo ARN y el dsRBD1 de ADAR2 (pdb 2L3C) y b) del motivo de reconocimiento bipartito del surco mayor del complejo ARNasa III-ARN (pdb 2NUE) ²¹².

el surco menor del ARNdh. Análisis de las estructuras de complejos de dsRBDs con ARNdh, muestran que la superficie de interacción de la hélice $\alpha 1$ con el ARN está constituida por cuatro a cinco aminoácidos que siempre se encuentran en las posiciones 3, 4, 7, 8 y 11 según el alineamiento de la figura 7. Es importante notar que, entre estos residuos, sólo algunos son parte de la secuencia consenso del dsRBD (posición 8). La falta de conservación de estas cadenas laterales podrían ser uno de los factores que modulan las propiedades de unión al sustrato de los diferentes dsRBDs ¹³⁹.

La segunda región del dsRBD que conforma la superficie de interacción con el surco menor del ARN es la que conecta las láminas $\beta 1$ y $\beta 2$. En el modelo de unión canónico, este bucle se inserta en el surco menor del ARN una vuelta de hélice más lejos de la hebra del surco menor que contacta la hélice 1. Este bucle tiene una longitud bien definida de 6 residuos, ubicados en las posiciones 28-33 en el alineamiento de la figura 7 y una composición de aminoácidos que sigue el patrón GPxHxx. El modo en que el bucle 2 se une al sustrato es por un enlace de hidrógeno entre el grupo carbonilo del esqueleto peptídico de la His 31 conservada

con el grupo 2'OH de una hebra, mientras que el grupo imidazol de este residuo se apila sobre una ribosa y forma un puente de hidrógeno con el 2'OH de la ribosa anterior de la otra hebra Figura 12. La conservación del modo de unión por el bucle 2 resulta de la conformación que adopta el esqueleto peptídico, donde la cadena lateral de la His 31 hace contactos de van der Waals con la cadena lateral de la Pro conservada (posición 29) y con el residuo ubicado en la posición 32, que generalmente es un grupo alifático. La cadena lateral del residuo 33, que generalmente es una lisina o una arginina, se pliega del lado del bucle que se apoya sobre el surco menor del ARN. En conjunto, estas interacciones forman un pequeño bolsillo hidrofóbico que tiene un efecto estabilizador para la conformación del bucle 2. La función central de la histidina 31 para la unión al ARN ha sido demostrada por estudios mutacionales ¹³⁶.

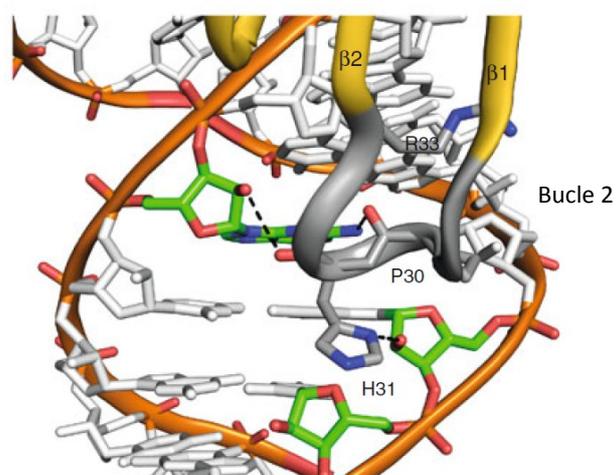


Figura 12: Reconocimiento del surco menor del ARNdh por el bucle 2 del dsRBD2 de XLRBPA (pdb 1DI2). Las cadenas laterales de los aminoácidos importantes para la unión al ARN están representados como palos ²¹².

1.6 FUNCIONES DE LOS dsRBDs EN LA BIOGÉNESIS DE miARNs

Los dsRBD son en principio capaces de unirse a cualquier elemento doble hebra en una molécula de ARN. Sin embargo, la presencia de imperfecciones en el ARNdh o la interacción con otras proteínas u otros dominios, le permiten tener especificidad sobre las regiones que reconocen ¹⁰⁹. Esta característica resulta especialmente apropiada para el reconocimiento de precursores de miARN,

debido a que estos no comparten similitudes en su secuencia y presentan un alto nivel de heterogeneidad. Si bien hasta el momento no se ha establecido una función clara de estos dominios durante la biogénesis de miARN, se ha demostrado que algunos dominios dsRBD presentan actividad chaperona de ARN, facilitando el apareamiento correcto de segmentos de ARN o promoviendo la formación de estructuras doble hebra en fragmentos que contienen protuberancias ^{123,124}. Por esta razón, una función posible para algunos de los dsRBDs involucrados en el procesamiento de miARNs en plantas, podría ser la de estabilizar elementos de estructura secundaria para permitir el posicionamiento correcto de los sitios activos de DCL1 en el complejo de procesamiento. Esta interacción daría lugar a la formación de posibles determinantes de la especificidad en la estructura de los precursores que no es aparente en las predicciones de estructura secundaria realizadas por programas estándar, y permitiría definir sin ambigüedades el registro de procesamiento.

1.6.1 dsRBDs DE DCL1

La presencia de los dominios dsRBD en las enzimas ARNasa III es esencial, sin embargo, sus funciones son variables. El único dominio dsRBD que tiene la ARNasa de *Aquifex aelicus* juega un rol principal en el reconocimiento de sustrato ¹⁵⁵. En las proteínas, la unión a sustrato es compartida entre sus dsRBDs y otros dominios con capacidad de unir ARN ⁵⁵. En DICER de humanos, el dsRBD está localizado al lado de la hendidura del sitio activo, sugiriendo su participación en la unión al ARNdh ⁶⁰. Recientemente, Ma y colaboradores mostraron que el dsRBD de DICER humana participa en el reconocimiento de sustrato junto con el dominio PAZ, pero que éste no es esencial para su actividad o para la determinación de la longitud del producto final ⁵⁵. El dsRBD de DICER de *Schizosaccharomyces pombe* no es necesario para la actividad ARNasa, pero se ha demostrado que participa en la localización subcelular de la proteína. Este

dominio, tiene una estructura no canónica, mostrando un bucle $\beta 1$ - $\beta 2$ de mayor tamaño y un sitio de unión a zinc(II). Por otro lado, este dominio es capaz de unir tanto ARNdh como ADN.

Una característica distintiva de las DCL de plantas es que DCL1, 3 y 4 presentan un segundo dsRBD en el extremo C-terminal (DCL1-B). Este dominio difiere en secuencia respecto al primero, que está más relacionado con el dsRBD de proteínas DICER de animales, y se sugiere que es un buen candidato para regular la especificidad de sustrato y/o la interacción con proteínas accesorias del complejo de procesamiento. Hay evidencia experimental que apoya ambas funciones. Se ha demostrado que DCL1-B participa en la determinación de la precisión de procesamiento *in vivo*³⁷. Esto sugiere que DCL1-B puede reconocer específicamente la estructura de los precursores asegurando el posicionamiento del complejo de procesamiento para liberar el miARN. Por otro lado, DCL1-B sería esencial para la localización de DCL1 en el sitio correcto de procesamiento de miARN en el núcleo, los cuerpos D. Sin embargo, no existe hasta el momento una caracterización detallada de este dominio. Las proteínas DCL2 son distintas de los otros tipos de DICER debido a que carecen del segundo dominio de unión a ARN doble hebra y con excepción de las diferencias que existen entre los dsRBD-A de DCL1 y DCL3, la variación neta que existe entre las combinaciones de pares de los tipos de DICER 1, 3 y 4 es mucho mayor en el dsRBD-B¹⁵⁶, lo que se podría relacionar con las funciones diferenciales que llevan a cabo cada una de estas proteínas.

Respecto a su función *in vivo*, se ha observado que mutaciones en DCL1-B, producen mutantes con pérdida de función, mientras que mutaciones que afectan al primer dsRBD de DCL1 son letales para el desarrollo del embrión. Las proteínas DICER en general realizan cortes de una distancia de 22 nucleótidos del extremo de ARN doble hebra, probablemente usando la distancia entre el

extremo del dominio PAZ y el centro catalítico de los dominios ARNasa III. En el procesamiento de miARN en humanos, la localización del miARN en el precursor está definida por el primer corte realizado en el núcleo por Drosha/DGCR8. En cambio en plantas DCL1 tiene que realizar el primer corte en el procesamiento y su segundo dsRBD podría ser crítico para definir la posición de este primer corte. Se ha probado que mutantes *dcl1-9* (presentan un T-DNA insertado en DCL1-B) dan cortes aberrantes del pri-miRNA163 ⁶¹.

La información actualmente disponible sobre el sistema se reduce a estudios bioquímicos sobre formas truncadas de DICER de otros organismos y a estudios genéticos sobre mutantes y es claramente insuficiente para explicar su mecanismo de acción.

1.6.2 dsRBDs DE HYL1

HYL1 es una proteína accesoria de DCL1. La participación de proteínas de unión a ARN doble hebra accesorias en la reacción es común a todas las maquinarias de procesamiento de ARNi caracterizadas hasta el momento. En animales, DROSHA es ayudada por la proteína de unión a ARNdH, DGCR8, mientras que el procesamiento de pre-miARNs por parte de DICER de mamíferos, requiere la presencia de la proteína TRBP ^{33,157}. En *D. melanogaster*, DCR-2 requiere la presencia R2D2 para llevar a cabo el silenciamiento de genes mediado por siARNs ¹⁵⁸. Todas estas proteínas accesorias presentan una arquitectura de dominios similar, que consiste en dos o tres dsRBDs organizados en tandem. Sin embargo, la función molecular exacta de estas proteínas que contienen solamente dsRBDs sin dominios catalíticos asociados al procesamiento es hasta el momento desconocida.

En *Drosophila melanogaster* la biogénesis de miARNs y siARNs requiere de maquinarias diferentes ¹⁵⁹. Los precursores de miARNs se procesan por los

complejos DROSHA/PASHA y DCR-1/LOQS, mientras que los ARNdh exógenos, son procesados por DCR-2/LOQS-R2D2 para generar siARNs. R2D2 y LOQS tienen funciones bioquímicas diferentes en el procesamiento mediado por las proteínas DICER. LOQS es requerida para el procesamiento de precursores de miARN, pero no es necesaria para cargar los miARN maduros en el complejo RISC^{31,160}. Por el contrario, R2D2 es necesaria para cargar los siARNs en el complejo RISC, pero no para el procesamiento de ARNdh exógenos^{158,161}. El gen *loqs* genera cuatro isoformas de ARNm por maduración alternativa: *loqs-RA*, *loqs-RB*, *loqs-RC*, y *loqs-RD*. Los cuatro ARNm producen cuatro proteínas distintas LOQS-PA, LOQS-PB, LOQS-PC, y LOQS-PD. Las distintas isoformas de estas proteínas, se asocian preferencialmente a DCR-1 o DCR-2 y de esta forma definen el procesamiento hacia la vía de miARNs o de siARNs^{30,31,162-164}. Por otro lado, durante la carga del complejo RISC, R2D2 orienta termodinámicamente los pares de siARN respecto a AGO2, permitiendo que se cargue la hebra correcta en el complejo efector y se degrade la otra^{158,165}. Además de participar en el ensamblado del complejo RISC, R2D2 dirige la actividad de DCR-2 incrementando la afinidad de la enzima para ARNdh largos con respecto a los pre-miARNs¹⁶⁶. En un trabajo realizado por Marques y colaboradores, se ha propuesto un mecanismo donde LOQS y R2D2 actúan secuencialmente durante la biogénesis de siARN donde LOQS promueve la actividad de DCR-2 sobre DCR-1 y R2D2 carga del complejo RISC de forma correcta¹⁶⁷. Recientemente, Fukunaga y colaboradores demostraron que LOQS-PB determina la localización de los sitios activos de DCR-1 sobre el pre-miR307a por un mecanismo hasta el momento desconocido. Los mismos efectos fueron observados para TRBP humana con el pre-miR-132¹⁶⁸.

Las propiedades de unión al ARN de estas proteínas accesorias son variables, por ejemplo, R2D2, contiene dos dsRBDs, sin embargo, no une precursores de siARN por sí sola, sino que necesita la presencia de DCR-2^{158,161,169}.

En TRBP, los primeros dos dominios interaccionan con el precursor de ARN, mientras que el tercero, es necesario para interaccionar con DICER¹⁷⁰⁻¹⁷². En HYL1, el primer dominio es el determinante en la unión al ARN, mientras que el segundo lo hace con menos afinidad^{76,77}. Aunque se supone que la unión al sustrato es parte de la función accesoria de estas proteínas, hasta el momento no ha sido establecido si es un requerimiento absoluto para la participación de las mismas en la biogénesis de miARNs, ni cuáles serían las consecuencias de alterar la afinidad de unión al sustrato en el mecanismo de procesamiento.

1.6.3 RECONOCIMIENTO DEL SITIO DE CORTE DE LOS PRECURSORES POR dsRBDs

La comprensión del mecanismo de reconocimiento de sustrato por parte de las proteínas de procesamiento de plantas, resulta de gran relevancia debido al papel que juega esta vía en el control y regulación de numerosos procesos esenciales. Si bien en los últimos tiempos ha existido un avance significativo en la caracterización de diversas proteínas encargadas del procesamiento de precursores de miARN de plantas, el actual grado de conocimiento sobre los detalles mecánicos y estructurales de estas proteínas es escaso. De hecho, hasta el momento no se ha podido esclarecer de manera clara que función desempeñan los dsRBDs localizados en el extremo C-terminal de DCL1, ni se han identificado cuáles son los dominios que posicionan al sustrato en el sitio activo de DCL1. De este modo, la información disponible hasta el momento no permite definir correctamente un modelo que permita explicar el mecanismo de reconocimiento específico de sustrato. Una caracterización exhaustiva de los dominios que interaccionan con el ARNdh sustrato, brindaría información valiosa respecto a la selección de sustrato y de esta forma daría un indicio de los motivos responsables de ensamblar la maquinaria de procesamiento de una manera coordinada y precisa en tiempo y espacio. Los resultados obtenidos no sólo

aportarían información respecto al mecanismo de silenciamiento génico en plantas, sino que también podrían aportar información valiosa respecto a la biofísica del reconocimiento molecular por parte de los dominios de unión a ARNdh.

2 OBJETIVOS

2.1 OBJETIVOS GENERALES

El objetivo general de este proyecto es la generación y validación de un modelo que permita explicar el reconocimiento de los sustratos por las proteínas DCL1 y HYL1 en el proceso de biogénesis de miARN. Con este fin estudiamos dominios de unión a ARN de la proteínas, por separado y en combinación, y su interacción con distintos precursores y variantes de los mismos.

2.2 OBJETIVOS ESPECÍFICOS

Durante el desarrollo del presente proyecto estudiamos los mecanismos de reconocimiento de los precursores de miARN por la enzima DCL1 y HYL1 e intentamos avanzar hacia la comprensión de las reacciones de procesamiento propiamente dichas.

Nuestra hipótesis es que DCL1 y HYL1 son capaces de reconocer la estructura terciaria de sus sustratos y situar los sitios catalíticos en una posición particular de esta estructura gracias a la especificidad de los dominios dsRBD. Para demostrar esta hipótesis nos propusimos cumplir los siguientes objetivos específicos.

- Caracterización estructural de los dominios dsRBD de las proteínas DCL1 y HYL1.
- Determinación de las afinidades de interacción con el sustrato ARNdh
- Estudio de los determinantes estructurales de la unión a sustrato de los dsRBDs de DCL1 y HYL1
- Análisis funcionales *in vivo* de estos dominios y variantes de los mismos.

3 MATERIALES Y MÉTODOS

3.1 CEPAS, PLÁSMIDOS Y MEDIOS DE CULTIVO

3.1.1 CEPAS BACTERIANAS

***Escherichia coli* BL21 (DE3):** genotipo $F^- ompT gal dcm lon hsdS_B(r_B^- m_B^-)$ $\lambda(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])$. El profago λ (DE3) incluye el gen que codifica para la ARN polimerasa del fago T7 bajo control del promotor lacUV5, y el gen lacI, que codifica para el represor de la transcripción LacI. Esta cepa se utilizó para la expresión de proteínas recombinantes.

***Escherichia coli* DH5 α** ¹⁷³: Esta cepa se usó para guardar los plásmidos obtenidos en un cepario. Su genotipo es: $fhuA2 \Delta (argF-lacZ) U169 phoA glnV44 \Phi80 \Delta(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17$.

***Escherichia coli* TOP10** (Invitrogen): Esta cepa se utilizó para realizar subclonados. Su genotipo es: $F^- mcrA \Delta(mrr-hsdRMS-mcrBC) \phi80lacZ\Delta M15 \Delta lacX74 nupG recA1 araD139 \Delta(ara-leu)7697 galE15 galk16 rpsL(Str^R) endA1 \lambda^-$.

***Agrobacterium tumefaciens* ASE:** Se utilizó para transformar *Arabidopsis thaliana*. Esta cepa es resistente a los antibióticos cloramfenicol y kanamicina.

3.1.2 VECTORES PLASMÍDICOS

Plásmido pET-TEV: Se utilizó como vector de expresión de proteínas recombinantes. Es un plásmido derivado del pET28-a al cual se le eliminó el sitio de corte de trombina y se reemplazó por el sitio de corte de la proteasa viral TEV. Presenta un sitio de iniciación de la transcripción para la T7 ARN Polimerasa, resistencia a kanamicina como marcador de selección y una secuencia codificante para una etiqueta de histidina.

Plásmido pUC-18: Se utilizó para subclonar proteínas recombinantes. Es un plásmido de alto número de copias. Tiene un gen de resistencia al antibiótico ampicilina y el fragmento N-terminal del gen de la β -galactosidasa de *E. coli* (*lacZ*) interrumpido por el sitio de múltiple clonado, lo que permite seleccionar las colonias que incorporan insertos en cepas que contienen la región C-terminal de la β -galactosidasa.

Plásmido pCHF3-CS052: El pCHF3 es un vector binario que permite la transformación de *A. thaliana* mediada por *A. tumefaciens*. Posee orígenes de replicación para *E. coli* y *A. tumefaciens*. También posee el gen de resistencia a espectinomicina para la selección bacteriana. El T-ADN está definido por los bordes derecho e izquierdo y es la porción de ADN transferida al genoma de la planta durante la transformación. Esta región contiene el gen *NPTII*, que confiere resistencia a kanamicina en plantas, bajo el promotor 35S del virus del mosaico del coliflor. Además presenta un sitio de múltiple clonado flanqueado por el promotor constitutivo 35S del virus del mosaico del coliflor y el terminador de la transcripción de la subunidad pequeña de Rubisco de arveja.

El vector pCHF3-CS052 es un pCHF3 clásico que tiene incorporada la secuencia correspondiente a la proteína verde fluorescente (GFP) de *Aequorea victoria*, lo que permite llevar a cabo la expresión de una proteína fusionada a esta proteína reportera.

Plásmido pCHF5: Vector binario que permite la transformación de *A. thaliana* mediada por *A. tumefaciens*. Posee orígenes de replicación para *E. coli* y *A. tumefaciens*. También posee el gen de resistencia a espectinomicina para la selección bacteriana. El T-ADN está definido por los bordes derecho e izquierdo y es la porción de ADN transferida al genoma de la planta durante la transformación. Esta región contiene el gen *BAR* que confiere resistencia al

herbicida glufosinato de amonio, bajo el promotor 35S del virus del mosaico del coliflor. Además presenta un sitio de múltiple clonado flanqueado por el promotor 35S del virus del mosaico del coliflor y el terminador de la transcripción de la subunidad pequeña de Rubisco de arveja.

3.1.3 MEDIOS DE CULTIVO

Para las purificaciones de proteínas se utilizaron dos medios de cultivo diferentes en función del tipo de experimento a realizar:

Expresión de proteína sin marca isotópica, clonados y pre-cultivos de expresión: Medio de cultivo rico Luria-Bertani (en adelante, LB) (Sigma): peptona de caseína 1% p/v, extracto de levadura 0,5% p/v, NaCl 0,5% p/v y kanamicina 50 µg/ml. En su versión sólida, este medio incluye agar-agar de 1 a 2,5% p/v.

Expresión de proteína marcada: Medio mínimo M9 (en adelante, M9): KH_2PO_4 25 mM, Na_2HPO_4 50 mM, NaCl 10 mM, MgCl_2 1 mM, CaCl_2 0,2 mM, Glucosa 0,4%, Solución de sales 1 ml/l (HCl concentrado 51,3 ml/l, MgCl_2 10,75 g/l, CaCO_3 2 g/l, $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 4,5 g/l, $\text{MnSO}_4 \cdot 4\text{H}_2\text{O}$ 1,12 g/l, $\text{CoSO}_4 \cdot 7\text{H}_2\text{O}$ 280 mg/l, $\text{H}_3\text{BO}_3 \cdot \text{H}_2\text{O}$ 60 mg/l, $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ 1,44 g/l) y 1 g/l NH_4Cl . Para la obtención de muestras con marcación isotópica se utilizó 1 g/l de $^{15}\text{NH}_4\text{Cl}$ o 1 g/l de $^{15}\text{NH}_4\text{Cl}$ y 2 g/l de $[\text{U-}^{13}\text{C}]$ glucosa (Cambridge Isotope Laboratories).

3.2 TRANSFORMACIÓN BACTERIANA

3.2.1 PRODUCCIÓN DE CÉLULAS COMPETENTES

Se generaron bacterias competentes de las cepas de *E. coli* descritas en el apartado 3.1.1. Para esto, se crecieron las células en 500 ml de LB hasta alcanzar una DO_{600} de $\sim 0,5$. Las células fueron enfriadas en hielo y cosechadas por centrifugación a 2600 g durante 10 min a 4°C. El sedimento celular se resuspendió en 100 ml de una solución fría de MgSO_4 100 mM estéril y se incubó

por 30 minutos en hielo. Posteriormente, las células fueron recuperadas por centrifugación a 2600 *g*, a 4°C durante 10 min y el sedimento obtenido se resuspendió en 10 ml de una solución fría y estéril de CaCl₂ 100 mM, glicerol 15%. Por último, se fraccionó el volumen total en muestras de 300 µl que fueron congeladas y conservadas a -80°C hasta el momento de la transformación.

3.2.2 TRANSFORMACIÓN DE *E.coli*

La transformación se realizó utilizando bacterias competentes generadas como se mencionó en el apartado anterior. Se tomaron alícuotas de 100 µl de bacterias competentes; se adicionó 1 µl de plásmido ó 5 µl de mezcla de ligación según corresponda, en el caso de los controles negativos se agregaron 1 ó 5 µl de agua como reemplazo del plásmido. Esta mezcla se incubó en hielo durante una hora, luego 90 segundos a 42°C y finalmente 2 min en hielo. A continuación se adicionaron 900 µl de medio LB y las células se incubaron durante una hora a 37°C. Transcurrida la incubación el cultivo celular se centrifugó 2 min a 2600 *g* a T° ambiente, el sedimento de células se resuspendió en 100 µl de medio LB y finalmente se sembró en placas de medio LB-agar preparadas con el antibiótico correspondiente. Las placas se incubaron toda la noche a 37°C.

3.2.3 TRANSFORMACIÓN DE *A. tumefaciens*.

Un cultivo de *A. tumefaciens* (cepa ASE), crecido toda la noche a 28-30°C en medio LB, fue diluido 1:100 en el mismo medio y cultivado hasta una DO₆₀₀ entre 0,5 y 1. La suspensión de células se fraccionó en alícuotas de 1,5 ml, cada una de las cuales sirve para realizar una transformación. Se centrifugó a 10.000 *g* por 1 min a 4°C, y se resuspendió el sedimento de células en 0,5 ml de agua destilada estéril a 4°C. Los lavados se repitieron 3 veces. A continuación, se resuspendieron las células en 0,5 ml de 10 % (v/v) glicerol estéril a 4°C. Se centrifugó nuevamente a 10.000 *g* por 1 min a 4°C y las células se

resuspendieron en 20 μ l de 10 % (v/v) glicerol estéril. La suspensión celular se transfirió a una cubeta de electroporación (Biorad) y se aplicó un pulso de 2,5 kV con capacitancia de 25 μ F y una resistencia de 400 Ω . A continuación se adicionó 1ml de medio LB y se incubó durante dos horas a 28°C. Transcurrida la incubación se realizó una centrifugación de 2 min a 2600 g a T° ambiente, el sedimento de células fue resuspendido en 100 μ l de medio LB y finalmente se sembró en placas de medio LB que contenían el antibiótico correspondiente. Las placas se incubaron toda la noche a 28°C.

3.2.4 SELECCIÓN Y ANÁLISIS DE TRANSFORMANTES.

La selección de las colonias *E. coli* transformadas se realizó suplementando las placas con medio LB con el antibiótico adecuado según el vector usado. Cuando el vector fue pUC-18, las placas se suplementaron 100 μ g/ml de ampicilina, para el vector pET-TEV se suplementaron con 50 μ g/ml de kanamicina y cuando los vectores fueron pCHF3 o pCHF5 con 80 μ g/ml de espectinomicina.

Además, en la selección de colonias transformadas con el vector pUC-18 se adicionó a las placas 100 μ M IPTG y 20 μ g/ml X-Gal. Esto permitió identificar potenciales colonias que contenían el vector con el inserto clonado por su ausencia de coloración azul.

Para la selección de las colonias ASE transformadas con los vectores binarios pCHF3-CS052 o pCHF5 las placas con medio LB se suplementaron con 50 μ g/ml de espectinomicina, 50 μ g de kanamicina y 25 μ g de cloramfenicol.

3.3 SUBCLONADO DE FRAGMENTOS DE ADN

3.3.1 MINIPREPARACIONES DE ADN PLASMÍDICO

Los plásmidos se purificaron utilizando "kit" de purificación de ADN

Wizard® SV Gel (Promega), siguiendo las indicaciones del fabricante.

3.3.2 ESTIMACIÓN DE LA CONCENTRACIÓN DE ADN

La concentración de ácidos nucleicos fue determinada a partir de espectros de absorción entre 340 y 240 nm en un espectrofotómetro Jasco 530 con cubetas de paso óptico de 1 cm. Los espectros fueron procesados con el programa Jasco Spectra Manager. La concentración se estimó asumiendo que 1 unidad de absorbancia a 260 nm de ADN corresponde a una concentración de 40 µg/ml.

3.3.3 AMPLIFICACIÓN DE FRAGMENTOS DE ADN MEDIANTE REACCIÓN EN CADENA DE LA POLIMERASA (PCR)

La amplificación de las secuencias de los dsRBDs de DCL1 se realizó a partir de un vector de expresión pET-TEV construido por el Dr. Nicolás Bologna en el laboratorio del Dr. Javier Palatnik. Este vector contiene la secuencia correspondiente a ambos dsRBDs de DCL1 (DCL1-AB). Para amplificar las secuencias que corresponden a los dsRBDs de DCL1 (DCL1-A y DCL1-B), se diseñaron oligonucleótidos que hibridan con las secuencia de interés e incluyen los sitios de restricción BamHI y Sall (indicados en letras minúsculas) necesarios para el clonado en el vector pCHF3-CS052.

DCL1-A Fw: GCGCg_gatccATGCATCCGGTGCAGAGCTACA

DCL1-A Rv: GCGCg_tcgacATTACCGTTGTTGATATGCTTCTC

DCL1-B Fw: GCGCg_gatccAATGGGCATCAGCCGTTTACGAG

DCL1-B Rv: GCGCg_tcgacAGTTTTATTTAAAAGCTCAAGTAG

Las reacciones de amplificación se llevaron a cabo en un volumen final de 50 µl. Se utilizaron 5 µl de tampón *Pfu* ADN polimerasa con MgSO₄ 10x

(Promega), 0,4 μ M de cada oligonucleótido, 0,1 μ g de ADN plasmídico y 1 U de *Pfu* ADN polimerasa. Se utilizó una concentración 0,2 mM de cada uno de los dNTPs (dCTP, dATP, dGTP, dTTP) y se completó el a volumen final con agua Milli-Q estéril.

Para la reacción de amplificación se utilizó el termociclador Perkin Elmer Gene Amp Pcr System 2400 con el siguiente programa:

Etapa	Ciclos	Temperatura	Tiempo
1	1	95	5 min
2	30	95	30 seg
		55	30 seg
		68	1 min
3	1	68	4 min

3.3.4 ELECTROFORESIS DE ADN EN GELES DE AGAROSA

La separación de moléculas de ADN por electroforesis en geles de agarosa se realizó mediante el sistema tipo submarino ¹⁷⁴ en cubas de acrílico. Se utilizaron concentraciones de agarosa de 1 a 2,5% (p/v) según el tamaño de las moléculas de ADN a resolver. Los geles se prepararon con solución amortiguadora TAE 1X (Tris-HCl 40 mM, Ácido acético glacial 20 mM y EDTA 1 mM) y se agregó el colorante de ADN SYBR® Safe (Invitrogen). Las muestras de ADN, previo a la siembra, se mezclaron con 1/6 volúmenes de solución de siembra (glicerol 30%, Xilene cyanol 0,25% azul de bromo fenol 0,25%). Las corridas electroforéticas se llevaron a cabo en solución amortiguadora TAE 1X, empleando voltaje constante. Se utilizaron marcadores de tamaño molecular adecuados.

3.3.5 PURIFICACIÓN DE FRAGMENTOS DE ADN A PARTIR DE GELES DE AGAROSA

Luego de la separación electroforética, el fragmento de ADN deseado se identificó mediante la visualización del gel con un transiluminador UV y se escindió la banda correspondiente con bisturí. La purificación de los fragmentos

de ADN del gel se llevó a cabo utilizando el "kit" de purificación AxyPrep DNA Gel Extraction Kit (Axygen).

3.3.6 DIGESTIÓN DE ADN CON ENDONUCLEASAS DE RESTRICCIÓN

Las digestiones de ADN con endonucleasas de restricción se realizó utilizando enzimas y tampones comerciales. Las condiciones de reacción fueron las especificadas por el fabricante.

3.3.7 LIGACIÓN DE FRAGMENTOS DE ADN

La ligación de fragmentos de ADN generados por corte con enzimas de restricción en los vectores plasmídicos correspondientes se llevó a cabo con la enzima T4 ADN ligasa (Fermentas). Las condiciones de la reacción fueron las especificadas por el fabricante. En todos los casos el volumen de reacción de ligación fue de 10 μ l y se utilizaron 5 μ l de la misma para transformar *E. coli* TOP10.

3.3.8 CHEQUEO DE TRANSFORMANTES

3.3.8.1 PCR DE COLONIAS

Se realizó con el fin de verificar la presencia del vector con el inserto en las colonias blancas seleccionadas al azar. Se resuspendió una porción de una colonia blanca en 50 μ l de agua desionizada. Se calentó 5 min a 95°C, se colocó en hielo y luego se centrifugó a 6.000 *g* por 2 minutos. Se descartó el pellet y se utilizaron 5 μ l de sobrenadante como fuente de molde para una reacción de PCR con las mismas características descritas en el apartado 3.3.3 excepto que se utilizó la enzima Taq polimerasa (Fermentas) con su solución amortiguadora y que la etapa de elongación se realizó 72°C.

3.3.8.2 ANÁLISIS DE RESTRICCIÓN

Para chequear la presencia de inserto de tamaño esperado, se realizó el corte con BamHI y Sall. Los cortes fueron realizados de acuerdo al protocolo descrito por el proveedor de la enzima y se analizaron mediante electroforesis en geles de agarosa 1% (p/v) Los tamaños de los fragmentos de ADN se estimaron utilizando como marcador ADN el marcador de peso molecular 100 pb Ladder (Promega).

3.3.8.3 SECUENCIACIÓN DE ADN

La secuenciación de los insertos realizó en el servicio de secuenciación de la Universidad de Maine (USA). Una vez obtenida la secuencia, se verificó la identidad de la misma utilizando el programa de acceso libre Ape (A Plasmid Editor) (<http://biologylabs.utah.edu/jorgensen/wayned/ape/>).

3.4 EXPRESIÓN Y PURIFICACIÓN DE PROTEÍNAS

3.4.1 EXPRESIÓN DE PROTEÍNAS RECOMBINANTES

Las secuencias correspondientes a las proteínas recombinantes fueron clonadas en un vector de expresión pET-TEV¹⁷⁵. Los plásmidos de expresión para DCL1 fueron construídos por el Dr. Nicolás Bologna en el marco de una colaboración con el grupo del Dr. Javier Palatink y los de HYL1 fueron subclonados en pET-TEV por el Lic. Fernando Milia como parte de su trabajo de tesina. En todos los casos, los dominios fueron definidos a partir de la herramienta PROSITE (<http://prosite.expasy.org/>). En el caso de DCL1-B, la construcción incluye los residuos N1826–S1909 de la proteína DCL1 con cinco residuos adicionales en el extremo N-terminal. El doble dominio DCL1-AB incluye los residuos H1733-S1909 con cinco residuos adicionales en el extremo N-terminal. En el caso de HYL1-A, la construcción incluye los residuos M1-K103,

mientras que HYL1-B está conformada por los residuos L101-S170 de la secuencia de HYL1, ambas construcciones incluyen 3 residuos extra en el extremo N-terminal. El doble dominio de HYL1 está diseñado desde M1-S170 con tres residuos adicionales en el extremo N-terminal.

A partir de una placa fresca de transformación de células de *E. coli* BL21 (DE3), se inoculó una colonia aislada en 2 ml de medio LB y se incubó durante 6 horas a 37°C con agitación constante a 220 rpm. Del pre-cultivo anterior se inocularon 500 µl en 50 ml de medio M9 (para proteínas con marca isotópica) y LB (para proteínas sin marca) y se incubó por 12 horas a 37°C con agitación constante. Finalmente, los 50 ml de este cultivo se trasvasaron a 1 l de medio M9 suplementado con 1 g/l de $^{15}\text{NH}_4\text{Cl}$ o 1 g/l de $^{15}\text{NH}_4\text{Cl}$ y 2 g/l de $[\text{U-}^{13}\text{C}]$ glucosa (Cambridge Isotope Laboratories) o LB, según corresponda. Las células fueron mantenidas a 37°C hasta alcanzar una OD_{600} de $\sim 0,6$. En este punto, se indujo la expresión de las proteínas recombinantes con IPTG 1 mM y se dejó crecer el cultivo durante 4 horas a 37°C.

3.4.2 PURIFICACIÓN DE PROTEÍNAS EN CONDICIONES NATIVAS

La purificación los dsRBDs de DCL1 se llevó a cabo en condiciones nativas utilizando las siguientes soluciones:

A: Tris 50 mM (pH 8), NaCl 500 mM, Imidazol 5 mM y β -mercaptoetanol 1 mM

B: Tris 50 mM (pH 8), NaCl 500 mM, Imidazol 25 mM y β -mercaptoetanol 1 mM

C: Tris 50 mM (pH 8), NaCl 500 mM, Imidazol 350 mM y β -mercaptoetanol 1 mM

D: Tris 50 mM (pH 8), NaCl 500 mM y β -mercaptoetanol 1 mM

E: Tris 50 mM (pH 8), NaCl 50 mM y β -mercaptoetanol 1 mM

Las células inducidas se cosecharon mediante centrifugación a 6000 *g* por 10

minutos. El sedimento celular se resuspendió en 20 ml de la solución A y la lisis de las células se realizó utilizando un sonicador. Se aplicaron 6 pulsos de 30 segundos con un intervalo de 1 minuto entre cada pulso. Con el objetivo de recuperar las proteínas en solución, el extracto de lisis se centrifugó a 2000 rpm en rotor ss-34 (Sorvall) 1 hora. El sobrenadante obtenido se sembró en una columna que contenía 2 ml de resina de "Sepharosa Ni-NTA superflow" (QUIAGEN), previamente equilibrada con 5 volúmenes de resina de **solución A**. Los lavados de la resina se realizaron con 20 ml de **solución B** para eliminar las proteínas unidas inespecíficamente. Por último la proteína de interés se eluyó con **solución C**. Las fracciones eluidas que contenían proteína fueron dializadas contra **solución D** para eliminar el imidazol y simultáneamente digeridas con proteasa TEV con etiqueta de Histidina¹⁷⁶. La proteasa fue eliminada en una columna de afinidad de Ni(II) y luego de dializarla con **solución E**, la proteína fue separada de los ácidos nucleicos por una cromatografía de intercambio iónico con una columna de "Sephadex CM" (GE). La columna se equilibró con **solución E** y se eluyó utilizando un gradiente de 50 mM a 1 M NaCl.

En algunos casos, las muestras fueron liofilizadas. Para esto, se realizó una diálisis a 4°C durante toda la noche contra una solución de Acetato de amonio 20 mM pH 6,5 y a continuación se colocaron en un liofilizador hasta la completa eliminación del agua.

3.4.3 PURIFICACIÓN DE PROTEÍNAS EN CONDICIONES DESNATURALIZANTES

Para las proteínas HYL1, se realizó una purificación en condiciones desnaturizantes para lo cual se utilizaron las siguientes soluciones:

I: PO_4^{-3} 100mM, Tris-HCl 10 mM, urea 8 M, pH 8,0

II: PO_4^{-3} 100mM, Tris-HCl 10 mM, urea 8 M, pH 6,3

III: PO_4^{-3} 100mM, Tris-HCl 10mM, urea 8 M, pH 4,5

La cosecha de células y la lisis celular se realizó en las mismas condiciones

que en el apartado anterior, a excepción que en este caso, el sedimento celular se resuspendió en 20 ml de la **solución I**. El sobrenadante obtenido por la centrifugación a 20000 rpm en rotor ss-34 (Sorvall) fue purificado utilizando una columna de afinidad de Ni(II) previamente equilibrada con 5 volúmenes de columna de **solución I**. Con el objetivo de eliminar proteínas unidas de manera inespecífica a la columna, se realizó un lavado con 20 ml de **solución II**. Por último la proteína de interés se eluyó con **solución III**, en fracciones de 1 ml.

3.4.4 REPLEGADO DE PROTEÍNAS

Para el plegado de las proteínas se optó por el método de replegado por diálisis. Para ello se dializaron las fracciones que contenían proteína contra 100 volúmenes de solución de replegado (PO_4^{3-} 100 mM (pH=7.0), NaCl 50 mM, β -mercaptoetanol 1 mM, Arg 50 mM, Glu 50 mM) por 12 horas. Todos los pasos de diálisis se realizaron utilizando membranas de diálisis de corte 10 kDa pre-tratadas por ebullición durante una hora en agua destilada y EDTA para remover las trazas de metales pesados.

Luego de dializar las proteínas se las concentró utilizando concentradores Amicon Ultra de corte 3 kDa, centrifugando a 6000 *g* a 4°C.

3.4.5 PURIFICACIÓN DE LA PROTEASA TEV

Para la purificación de la proteasa TEV se utilizaron las siguientes soluciones:

O: Tris 50 mM (pH 8), NaCl 500 mM, Glicerol 10 %, Imidazol 20 mM y β -mercaptoetanol 1 mM

P: Tris 50 mM (pH 8), NaCl 1,5 M, Glicerol 10 %, Imidazol 100 mM y β -mercaptoetanol 1 mM

Q: Tris 50 mM (pH 8), NaCl 200 mM, Glicerol 10 %, EDTA 2 mM y β -mercaptoetanol 1 mM.

Las células inducidas se cosecharon mediante centrifugación a 6000 *g* por 10 minutos. El sedimento celular se resuspendió en 25 ml de la **solución O** y a continuación se sonicaron las células de la misma forma que en el apartado 3.4.2. El sobrenadante se sembró en una resina de afinidad de Ni (II) y se realizó un lavado con la **solución P**. La elución se llevó a cabo con un gradiente de Imidazol de 100 a 500 mM. A continuación, se juntaron las fracciones que contenían proteína y se agregó DTT y EDTA 1mM cada uno y se dializó contra la **solución Q** durante toda la noche a 4°C. Finalmente, se adicionó Glicerol 10 % y DTT 10 mM, se fraccionó y se almacenó en un freezer -80°C.

La evaluación de la actividad de la proteasa se realizó utilizando proteína HYL1-1 fusionada a una etiqueta de histidina con sitio de corte de TEV. Se realizaron cortes de distinta relación proteasa:proteína de fusión, desde 1:300 hasta 1:12. La digestión se realizó a 4°C durante toda la noche y la evaluación de la actividad se observó a partir de geles de poliacrilamida 15%.

3.4.6 ESTIMACIÓN DE LA CONCENTRACIÓN DE PROTEÍNAS

La concentración de proteínas fue estimada con espectros de absorción entre 340 y 240 nm en un espectrofotómetro *Jasco v-530* con cuβs de paso óptico 1 cm. Los espectros se procesaron en el programa *Jasco Spectra Manager*. En cada caso, se utilizó un coeficiente de absorción a 280 nm obtenido a partir del programa "protparam" (<http://www.expasy.org/tools/protparam.html>), utilizando la secuencia primaria de cada proteína.

3.4.7 ELECTROFORESIS DE PROTEÍNAS EN GELES DE POLIACRILAMIDA

La electroforesis de proteínas en geles de poliacrilamida se realizó en

condiciones desnaturalizantes y reductoras. Se utilizó un sistema discontinuo que consiste en un gel de concentración (poliacrilamida 4,5% en Tris-HCl 0,126 M, pH 6,8, SDS 0,26% p/v) seguido por un gel de separación (poliacrilamida 15% en Tris-HCl 0,36 M, pH 8,8, SDS 0,26% p/v). A las muestras se les adicionó la solución de siembra (glicerol 5%, SDS 2%, β -mercaptoetanol 0,1%, azul de bromofenol 0,1 mg/ml) y posteriormente se calentaron durante 5 minutos a 100°C. Para la corrida se utilizó una solución amortiguadora de Tris 0,3% p/v, glicina 1,44% p/v y SDS 0,1% p/v y se aplicó una corriente constante de 30 mA por gel. En todos los casos se empleó el sistema Miniprotean 4 (Bio-Rad Laboratories). Luego de la corrida, las proteínas se fijaron tratando el gel durante 5 minutos en una solución decolorante [etanol: ácido acético: agua (30:10:60)] con agitación y se tiñeron con una solución de Azul Brillante de Coomásie R250 1 % p/v en etanol: ácido acético: agua (50:10:40). La decoloración de los geles se consiguió manteniendo el gel en solución decolorante con agitación constante.

3.5 ESPECTROSCOPÍA DE RESONANCIA MAGNÉTICA NUCLEAR

3.5.1 FUNDAMENTOS TEÓRICOS

3.5.1.1 EL ESPECTRO $^1\text{H}^{15}\text{N}$ -HSQC

El experimento $^1\text{H}^{15}\text{N}$ -HSQC (*Heteronuclear Single Quantum Coherence*) es la base de gran parte de los estudios de RMN multidimensional en proteínas. En este experimento se correlaciona el desplazamiento químico de un núcleo de ^1H con el de un núcleo de ^{15}N al que se encuentra directamente unido, con una constante de acoplamiento escalar (1J) de 90–94 Hz. Así, en el espectro $^1\text{H}^{15}\text{N}$ -HSQC de una proteína se espera observar un número total de señales igual al número de residuos de la proteína, sin contar los residuos de prolina, que no poseen ^1H amídico y, por ende, no generan señal en este espectro. A las señales correspondientes a los grupos amida del esqueleto de la proteína se suman las

señales correspondientes a los grupos amida de las cadenas laterales de Gln y Asn y las señales de los pares NH de los anillos indólicos de los residuos de Trp. Teniendo en cuenta la información que aporta el espectro $^1\text{H}^{15}\text{N}$ -HSQC, el potencial de dicho experimento para los estudios biofísicos de proteínas depende de que se disponga de la asignación de sus señales. Es decir, para poder interpretar los cambios observados en los espectros adquiridos en estados diferentes de la proteína, es necesario determinar a qué residuo corresponde cada señal.

La posición de cada señal en el espectro depende fuertemente del microentorno químico del par $^1\text{H}^{15}\text{N}$ amida correspondiente. Esta dependencia convierte al espectro $^1\text{H}^{15}\text{N}$ -HSQC en la **huella dactilar** de una proteína bajo las condiciones de estudio. Por este motivo, el $^1\text{H}^{15}\text{N}$ -HSQC es el experimento ideal, en la gran mayoría de los casos, para estudiar la interacción de una proteína con ligandos, así como también el efecto del pH, la temperatura o la fuerza iónica, entre muchas otras aplicaciones. Otro experimento equivalente al $^1\text{H}^{15}\text{N}$ -HSQC que permite correlacionar los desplazamientos químicos de ^1H con los de un heteronúcleo es el $^1\text{H}^{15}\text{N}$ -HMQC (*Heteronuclear Multiple Quantum Coherence*). Este experimento brinda una mayor relación señal:ruido que el $^1\text{H}^{15}\text{N}$ -HSQC y puede ser utilizado en los casos en que la resolución no sea un punto crítico.

3.5.1.2 ASIGNACIÓN DE RESONANCIAS ^1H , ^{15}N Y ^{13}C

Para llevar a cabo la asignación de las señales de un $^1\text{H}^{15}\text{N}$ -HMQC, la estrategia mas utilizada consiste en la adquisición de experimentos de triple resonancia. En estos experimentos, cada par $^1\text{H}^{15}\text{N}$ es correlacionado con el valor de desplazamiento químico de distintos núcleos de ^{13}C del mismo residuo y/o de su residuo anterior en la secuencia proteica, según el caso. El éxito de esta estrategia se basa en que los desplazamientos químicos de ^{13}C de los carbonos $\text{C}\alpha$

y C β de un dado aminoácido son más dependientes de la identidad química del mismo que de su entorno.

Para resolver la estructura tridimensional de una proteína en solución, es necesario asignar a todos los protones de la misma, principalmente, aquellos de las cadenas laterales. Al igual que en el caso de los espectros de asignación del esqueleto peptídico, la estrategia para asignar cadenas laterales se basa en la adquisición de experimentos que permitan correlacionar los distintos núcleos de un residuo a través de constantes de acoplamiento escalares (a través de enlaces). Los espectros utilizados con tal fin generalmente son el H(CC)(CO)NH y (H)CCH-TOCSY. Este tipo de experimentos, permiten correlacionar los desplazamientos químicos ^{13}C y ^1H , respectivamente, de un sistema de espín completo.

Algunos de los experimentos utilizados para la asignación de las resonancias del esqueleto peptídico y de las cadenas laterales de proteínas se muestran en la figura 13.

3.5.1.3 CÁLCULO DE LA ESTRUCTURA EN SOLUCIÓN

Para la determinación de la estructura tridimensional de una proteína en solución es necesario que una vez asignadas las resonancias correspondientes a todos los protones de la proteína (que fuera posible asignar), se obtengan las restricciones estructurales de distancias, las constantes de acoplamiento escalar de tercer orden ($^3J_{\text{HN-HA}}$) y los ángulos que permitirán ubicar los distintos ^1H en el espacio de la proteína. Para obtener las restricciones de distancia se utilizan espectros 3D NOESY (*Nuclear Overhauser Effect Spectroscopy*) editados ^{15}N y ^{13}C . En este tipo de espectros las señales observadas corresponden a un par de ^1H cercanos en el espacio, acoplados dipolarmente entre sí. Dado que la intensidad de tales señales depende de la recíproca de la sexta potencia de la distancia

entre los dos ^1H , la integración de todas las señales la integración de todas las señales en un espectro NOESY deriva en una lista de restricciones de distancias ^1H - ^1H .

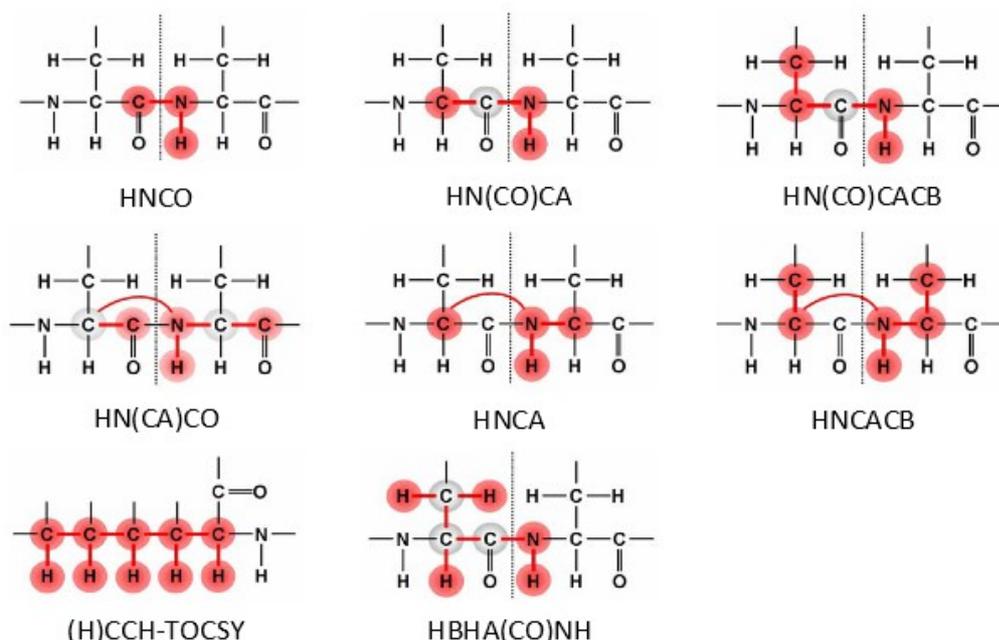


Figura 13: Esquemas de los experimentos de triple resonancia utilizados generalmente para asignar el esqueleto peptídico (HNCO, HN(CA)CO, HN(CO)CA, HNCA, HN(CO)CACB y HNCACB) y para los protones de las cadenas laterales ((H)CCH-TOCSY y HBHA(CO)NH). En rojo se indican los núcleos que son detectados y en gris se indican aquellos a través de los cuales se filtra la magnetización. Las líneas rojas indican los caminos de transferencia de la magnetización.

Las restricciones de ángulos diedros pueden obtenerse por medio de diversas estrategias. Las que se utilizan generalmente incluyen el cálculo de los ángulos diedros ϕ y ψ para cada residuo a partir de la evaluación de los valores de desplazamiento químico de los núcleos HN, N, CO, $\text{C}\alpha$, $\text{H}\alpha$ y $\text{C}\beta$ y la determinación de los valores de constante de acoplamiento de tercer orden entre los protones amida y alfa ($^3J_{\text{HN-H}\alpha}$). Las constantes de acoplamiento escalar $^3J_{\text{HN-H}\alpha}$ pueden determinarse a partir de un experimento HNHA midiendo la

relación de intensidades entre una señal y el pico diagonal correspondiente según la ecuación ¹⁷⁷:

$$\frac{I_{señal}}{I_{diagonal}} = -\tan^2(2\pi J_{HN-HA} \xi)$$

donde *Iseñal* e *Idiagonal* son las intensidades mencionadas y ξ es el tiempo de evolución del acoplamiento que se configura en el experimento (d23 en la secuencia **hnhagp3d** de Bruker).

Una vez determinado el valor de $^3J_{HN-HA}$ es posible obtener el valor del ángulo diedro phi (ϕ) a partir de la siguiente ecuación:

$$J_{HN-HA} = A \cos^2(\phi - 60) + B \cos(\phi - 60) + C$$

donde A, B y C son parámetros adimensionales optimizados empíricamente con valores A = 6,51; B = -1,76 y C = 1,60.

Las restricciones angulares y de distancias máximas obtenidas son utilizadas como entrada en un programa que las aplica sobre una familia de estructuras generadas al azar a partir de la secuencia de la proteína. El programa ajusta los ángulos de todos los enlaces de la macromolécula mediante un algoritmo de dinámica molecular que incluye un término de energía artificial que describe cuán bien se satisfacen las restricciones suministradas como datos de entrada. Las estructuras que mejor satisfacen las restricciones impuestas componen la salida del programa, junto con la lista de restricciones que resulten violadas en cada estructura de la familia. Finalmente el programa realiza un proceso de refinamiento en el que se revisan las restricciones violadas, se corrigen o se remueven de la lista de restricciones original, y se ejecuta nuevamente el programa hasta que las violaciones son mínimas.

Durante la determinación de estructuras tridimensionales, un paso que

consume mucho tiempo es la asignación e integración de las señales en los espectros NOESY. Por este motivo, se han desarrollado algoritmos que automatizan total o parcialmente los procesos de asignación inicial, asignación e integración de las señales de estos espectros. Uno de los programas que incluyen estos algoritmos es UNIO'10, que combina los programas ATNOS/CANDID, para realizar la asignación e integración de los espectros NOESY, y CYANA, que realiza el cálculo de estructura propiamente dicho mediante la aplicación de un algoritmo de *simulated annealing* con dinámica de ángulos diedros^{178,179}. En particular, UNIO'10 realiza siete ciclos iterativos de ATNOS/CANDID + CYANA, en cada uno de los cuales se obtiene una familia de estructuras compatibles en su mayor parte con las señales halladas y utiliza la información estructural derivada de ciclos anteriores para evaluar las señales con asignación ambigua. Específicamente, en cada ciclo el programa calcula 80 estructuras posibles y, de ellas, selecciona las 20 estructuras de menor energía (en relación al número y tipo de violaciones presentes en cada una) para ser usadas como molde de la iteración siguiente. La familia de estructuras finalmente obtenida incluye las 20 estructuras que mejor satisfacen las restricciones estructurales derivadas del análisis de los espectros de RMN. Otro programa que permite calcular estructuras de proteínas incorporando además restricciones de RDC (ver sección 3.5.1.4) es ARIA 2.3.1/CNS 1.1. Este programa realiza el cálculo de estructura de una forma similar a UNIO'10, pero además, en el último paso del cálculo incorpora un refinamiento final por dinámica molecular en el solvente seleccionado.

3.5.1.4 ACOPLAMIENTOS RESIDUALES DIPOLARES

Los parámetros de RMN que se utilizan generalmente para obtener información estructural en proteínas son los desplazamientos químicos (δ), las constantes de acoplamiento de tercer orden entre los protones amida y alfa (${}^3J_{\text{HN}}$).

$_{HA}$) y el NOE. Las medidas de acoplamientos dipolares residuales (RCDs) son un complemento más y en algunos casos pueden resultar cruciales para la determinación de estructuras de biomoléculas ¹⁸⁰. Esto se debe a que los RDC aportan información de largo alcance debido a que permiten conocer la orientación relativa entre diferentes pares de núcleos de la proteína. Si bien las restricciones de distancia que se obtienen a partir de la asignación de cadenas laterales y los NOE son de suma importancia en el cálculo estructuras de proteínas, se ha demostrado que también es posible obtener estructuras de proteínas de un tamaño hasta 200 residuos a partir de valores de desplazamiento químico y de RDCs ¹⁸¹. Los datos de RDC pueden ser incorporados a el programa CS-Rosetta que calcula estructuras de macromoléculas biológicas a partir de los valores de desplazamientos químicos de $^{13}C\alpha$, $^{13}C\beta$, ^{13}CO , ^{15}N , $^1H\alpha$ y 1HN y los datos de RDC. El programa realiza el cálculo utilizando valores de desplazamiento químico de fragmentos de proteínas seleccionados del Protein Data Bank.

3.5.1.5 DETERMINACIÓN DE LOS PARÁMETROS DINÁMICOS EN SOLUCIÓN

3.5.1.5.1 PARÁMETROS DE RELAJACIÓN

Numerosas funciones de las proteínas están ligadas a la movilidad interna de las mismas. Sin embargo la resolución de la estructura tridimensional de las proteínas por técnicas como cristalografía de rayos X y espectroscopía de RMN, nos da una imagen estática de ellas. Es igual de importante poder determinar cómo y cuánto se mueven las diferentes partes de ésta maquinaria, en la escala de tiempo apropiada para que realice su función, dado que nos permitiría lograr un mayor conocimiento y un mejor entendimiento de los distintos procesos en los que está involucrada.

Para analizar procesos dinámicos, que ocurren a escalas de tiempo

diferentes, se pueden utilizar distintas técnicas, pero RMN es la única técnica que nos permite hacer un seguimiento de los mismos con resolución atómica (Figura 14).

En espectroscopía, la relajación es el camino por el cual una partícula que se encuentra en un estado excitado vuelve a su estado basal. En RMN el retorno al equilibrio de los espines excitados se produce por la fluctuación de campos magnéticos locales causados por los movimientos moleculares. Estas fluctuaciones inducen transiciones de energía de espines y pérdida de coherencia de fase, llevando al sistema de regreso al equilibrio. Como la relajación se ve afectada por los movimientos moleculares, la velocidad de relajación medida por RMN contiene información dinámica de la molécula.

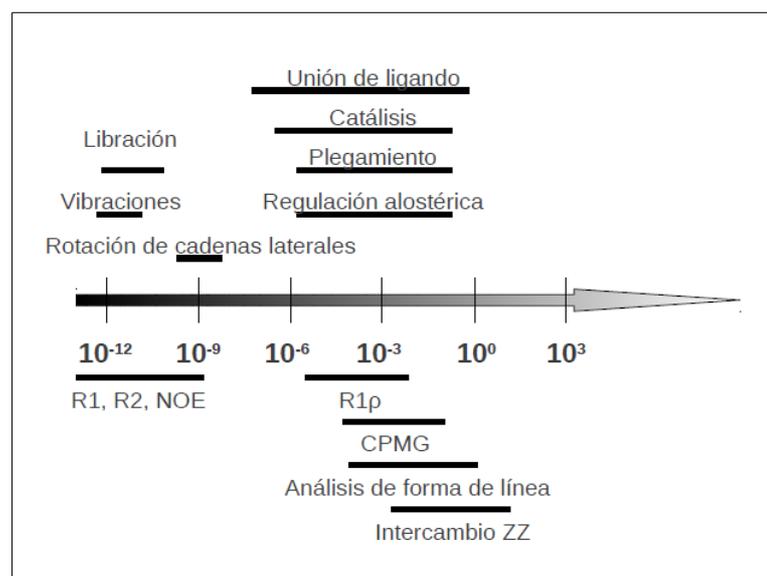


Figura 14: Escalas de tiempo de la dinámica de las proteínas y distintas técnicas de RMN que se pueden usar para estudiarlas.

Para poder entender el concepto de relajación de los espines nucleares es importante definir el estado de los mismos estando totalmente relajados en un campo magnético estático externo. En ésta condición la población de cada

estado de espín (a y b para $I=1/2$) sigue una distribución de Boltzman y no existe coherencia en la precesión de los espines. La aplicación de un pulso de radiación electromagnética a una frecuencia resonante, produce una distorsión de éste equilibrio: por ejemplo un pulso de 180° intercambia las poblaciones de espín rompiendo el primer equilibrio; un pulso de 90° produce coherencia de la magnetización transversal rompiendo la segunda característica del equilibrio ¹⁸².

La vuelta al equilibrio en RMN se produce en forma diferente dependiendo de la orientación de la magnetización en estudio respecto del campo magnético (Figura 15). En primer lugar, la relajación longitudinal, que permite la restauración de la distribución de Boltzman de los espines, es la vuelta al equilibrio de la componente de la magnetización paralela al campo. Ésta tiene su origen en el acoplamiento de los espines excitados con sus alrededores y se da mediante transferencia de energía, en forma de calor, al entorno del núcleo. Éste proceso está caracterizado por el tiempo de relajación longitudinal T_1 o la frecuencia R_1 ($1/T_1$).

Por otro lado, la relajación espín-espín o transversal, que corresponde a las componentes del vector de magnetización en orientación perpendicular al campo magnético, se define por el tiempo de relajación transversal T_2 o la frecuencia R_2 ($1/T_2$). Éste tipo de relajación es la causante del defasaje de la magnetización transversal mientras ésta precesa alrededor del campo estático externo (pérdida de coherencia). El T_2 surge de variaciones locales de campo magnético. Son tres mecanismos principales los que producen éste efecto: la inhomogeneidad del campo magnético, los acoplamientos dipolares y la anisotropía del desplazamiento químico.

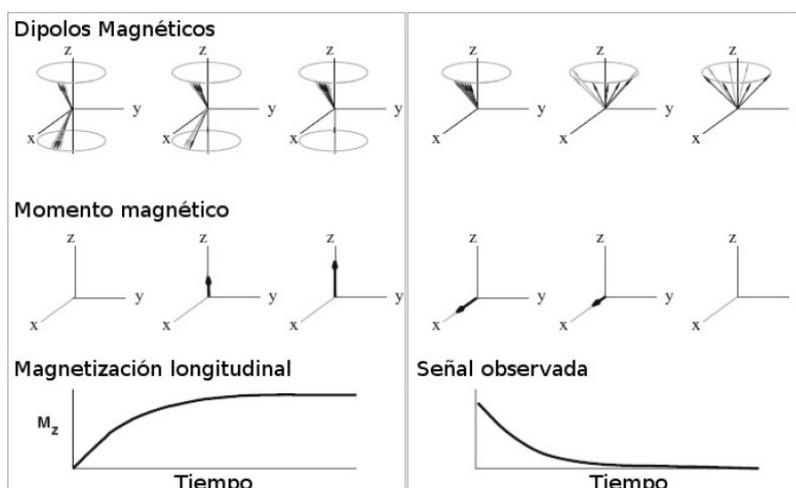


Figura 15: Relajación de la magnetización longitudinal (izquierda) y transversal (derecha). Se muestran la distribución de espines (arriba), el momento magnético resultante (centro) y la evolución de la magnetización relevante en el tiempo (abajo) (Rule & Hitchens 2006).

Otro parámetro de RMN que contiene información sobre la dinámica de proteínas es el NOE heteronuclear entre el ^{15}N y el ^1H unido. El valor del NOE heteronuclear refleja la relajación cruzada entre éstos núcleos, que a su vez depende de la frecuencia de movimiento molecular de la proteína.

Los parámetros R1 y R2 se obtienen a partir de un ajuste de las intensidades de las señales en función del tiempo de evolución correspondiente a cada espectro. Las curvas de relajación responden a una función exponencial simple. Los datos se ajustaron utilizando tres parámetros según la siguiente ecuación:

$$I(t) = I_0 e^{(-t/T_1)} + I_\infty$$

siendo I la intensidad en altura o volumen de la señal y t el tiempo de retardo utilizado en el espectro. Los parámetros que se obtienen del ajuste son I_0 que es la intensidad de la señal a tiempo de retardo 0, $T_{1/2}$ que es el tiempo de

relajación longitudinal o transversal I_{∞} que es la intensidad de la señal a tiempo infinito.

Para completar el conjunto de datos necesarios para el análisis, se deben obtener los valores de NOE ^1H - ^{15}N . Los experimentos de NOE consisten en dos espectros, uno en que se saturan los ^1H (saturado) y otro en que no se hace (referencia). El programa, extrae los valores de intensidad correspondientes a cada señal y se estima el NOE_{HN} de la siguiente manera:

$$\text{NOE}_{HN} = \frac{I_{\text{sat}}}{I_{\text{ref}}}$$

donde I_{sat} es la intensidad de la señal del espectro saturado e I_{ref} es la intensidad correspondiente al espectro de referencia.

El procesamiento de los datos de relajación, T_1 , T_2 y $\text{NOE}_{\text{HN-N}}$ se realizó utilizando las rutinas implementadas en el programa CCPNMR

3.5.1.5.2 DETERMINACIÓN DEL TENSOR DE DIFUSIÓN ROTACIONAL

Dado que la velocidad de relajación depende del tiempo de correlación rotacional de cada ^{15}N amida, y este se ve afectado principalmente por el movimiento global de la molécula, el primer paso en el análisis de relajación es determinar el valor de los tensores de difusión rotacional de la molécula. En el caso de una molécula esférica la movilidad es isotrópica y la función de correlación rotacional es una simple exponencial, pero en el caso de que la molécula sea asimétrica, la rotación también lo será y la función de correlación rotacional resulta multiexponencial. El ajuste de los datos asumiendo difusión isotropía será inexacto si la molécula en realidad presenta cierto grado de anisotropía. Por eso es importante la estimación del tensor de difusión rotacional de la molécula previo al análisis de la movilidad interna de la proteína. La difusión rotacional puede ser estimada utilizando la relación R_2/R_1 , debido a que

esta relación es independiente de la dinámica interna pero no del movimiento de la proteína como un todo. El programa TENSOR2, define si la movilidad de la proteína responde a un modelo isotrópico, axialmente simétrico o totalmente anisotrópico a partir de los valores de R1 y R2 ingresados. La elección del modelo se realiza mediante un análisis χ^2 de los datos a través de simulaciones Monte Carlo, definiendo el grado de confianza para el modelo que se analiza. El programa comienza evaluando el ajuste al modelo mas simple, modelo isotrópico, y luego aumenta en complejidad hasta el modelo totalmente anisotrópico. Para evaluar si el aumento en la complejidad del modelo es valido, se realiza un análisis estadístico F que evalúa si la mejora del ajuste es significativa como para rechazar el modelo mas simple. Para disminuir la probabilidad de que la movilidad interna afecte el calculo, el programa selecciona los valores de R1 y R2 de residuos que forman parte de estructuras secundarias definidas o comparando el resultado de la relación de cada ^{15}N amídico respecto a la media. Los valores propios del tensor de difusión reflejan la forma global de la proteína, si es esférica, prolato u oblato ¹⁸²⁻¹⁸⁶.

3.5.1.5.3 ANÁLISIS MODEL FREE

Una vez determinado el tiempo de correlación rotacional y el tensor de difusión el programa tiene el punto de partida para continuar con el análisis dinámico del esqueleto proteico.

Los parámetros de relajación R1, R2 y NOE son funciones de la densidad espectral $J(\omega)$ según se muestra en las siguientes ecuaciones:

$$R1 = d[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + cJ(\omega_n)$$

$$R2 = (d/2) [4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N) + 6J(\omega_H)] +$$

$$(c/6) [4J(0) + 3J(\omega_n)] + R_{ex}$$

$$\sigma_{NOE} = 1 + d (\gamma_H/\gamma_N) [6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]/R1$$

Donde $d = (\mu_0 \gamma_H \gamma_N h / 8\pi^2) \langle r_{NH}^{-3} \rangle$ y $c = [\omega_N (\sigma^{\parallel} - \sigma^{\perp})] / \sqrt{3}$, h es la constante de Planck, γ_H y γ_N son las constantes giromagnéticas de ^1H y ^{15}N respectivamente, σ^{\parallel} y σ^{\perp} son las componentes paralelas y perpendiculares al tensor de desplazamiento químico de ^{15}N y r_{NH} es la distancia entre el H y el N amídico. R_{ex} es la contribución de intercambio que se da cuando el ^{15}N se mueve entre distintos ambientes magnéticos en la escala de micro a milisegundos. La técnica de mapeo reducido de la densidad espectral nos permite convertir estos valores de R1, R2 y NOE en tres valores de densidad espectral $J(0)$, $J(\omega_N)$ y $J(\omega_H)$, considerando que como $\omega_H = 10\omega_N$ los valores de densidad espectral $J(\omega_H - \omega_N)$, $J(\omega_H + \omega_N)$ y $J(\omega_H)$ son todos aproximadamente iguales y se representan por $J(\omega_H)$. Esta aproximación reduce el problema a un ajuste simple de tres ecuaciones y tres incógnitas.

El análisis *model free* consiste en un análisis de minimización de χ^2 de los parámetros de relajación. El análisis se hace cinco veces, una vez por cada uno de los modelos que se explican a continuación, aumentando en complejidad, hasta seleccionar a cual de ellos se ajustan mejor los datos ingresados con un límite de confianza de 95%.

Los distintos modelos son:

1. Se sigue un ajuste simple *model free*, Asumiendo que τ_i es muy rápido ($\tau_i < 20\text{ps}$) la ecuación de densidad espectral puede expresarse de esta manera para el ajuste:

$$J(\omega) = \frac{2(S^2 \tau c)}{5(1 + \omega^2 \tau c)}$$

En este modelo el único parámetro ajustado es S^2 que indica la amplitud de movimiento del par HN amídico.

2. Sigue el modelo simple *model free* pero en este caso se considera que la

movilidad interna afecta a la función de correlación rotacional y la ecuación de densidad espectral queda expresada:

$$J(\omega) = \frac{2}{5} \left[S^2 \tau \frac{c}{(1 + \omega^2 \tau c)} + (1 - S^2) \tau \frac{e}{(1 + \omega^2 \tau e)} \right]$$

En este modelo se ajustan S^2 y τ_i , dado que $\tau e^{-1} = \tau c^{-1} + \tau i^{-1}$

3. Igual que modelo 1 con contribución de intercambio químico.

$$J(\omega) = \frac{2}{5} \frac{(S^2 \tau c)}{(1 + \omega^2 \tau c)} + Rex$$

Se ajustan S^2 y Rex .

4. Igual que modelo dos con contribución de intercambio químico. Se ajustan τ_i , S^2 y Rex .

5. Si el modelo simple no es suficiente para obtener un buen ajuste, se incluye un tiempo de correlación intermedio dentro de la función de densidad espectral, por lo que se usa lo que se llama sistema *model free* extendido. La ecuación es parametrizada por un tiempo de correlación rápido τ_f y un tiempo de correlación lento τ_s , así como también por dos parámetros de orden S_f^2 y S_s^2 correspondientes a esos tiempos. Siendo $\tau_f \ll \tau_s \ll \tau_c$, $S^2 = S_f^2 S_s^2$:

$$J(\omega) = \frac{2}{5} \left[S^2 \frac{(\tau c)}{(1 + \omega^2 \tau c^2)} + (1 - S_f^2) \frac{(\tau' f)}{(1 + \omega^2 \tau' f^2)} + (S_f^2 - S^2) \frac{(\tau' s)}{(1 + \omega^2 \tau' s^2)} \right]$$

Los parámetros $\tau_f^{-1} = \tau_c^{-1} + \tau_f^{-1}$ y $\tau_s^{-1} = \tau_c^{-1} + \tau_s^{-1}$. Como τ_f es muy rápido la ecuación se aproxima a:

$$J(\omega) = \frac{2}{5} \left[S^2 \frac{(\tau c)}{(1 + \omega^2 \tau c^2)} + (S_f^2 - S^2) \frac{(\tau' s)}{(1 + \omega^2 \tau' s^2)} \right]$$

Los ajustes de los datos de relajación en este modelo dan como resultado S_f^2 , S_s^2 y τ_s .

Por último para obtener el error de los ajustes se realiza un análisis de simulaciones Monte Carlo que evalúa las posibles variaciones de los parámetros calculados dentro del error experimental de los datos informados. En los casos en que ninguno de los modelos pueda reproducir los datos satisfactoriamente se lo considera modelo 6 y el programa proporciona los parámetros dinámicos obtenidos con el modelo de mejor ajuste. A través de este análisis el programa, teniendo en cuenta el tiempo de correlación rotacional de la molécula τ_c y los valores de los tensores de difusión de la molécula, nos permite obtener la movilidad interna de los mismos dada por el S^2 y un tiempo de correlación interno τ_i asociado al mismo, para los distintos pares HN amídico de los cuales se estiman los parámetros de relajación. En el caso de que los residuos que presentan intercambio rápido (μs a ms) entre dos ambientes químicos distintos, también se obtiene la constante de intercambio químico $R_{\text{ex}}^{182-186}$.

3.5.1.6 INTERCAMBIO QUÍMICO

Cuando un núcleo de una muestra puede encontrarse en dos entornos diferentes que resultan en distintos desplazamientos químicos, el espectro de RMN resultante dependerá de la relación entre la velocidad de intercambio entre los entornos y la diferencia de frecuencias de resonancia del núcleo en cada entorno. El fenómeno de intercambio puede darse en un sistema simple en el que ocurren cambios conformacionales, o durante una titulación de dos especies en donde las formas libre y unida de cada especie dan lugar a una frecuencia de resonancia diferente para el mismo núcleo.

Debido a las magnitudes de las diferencias de frecuencia que se encuentran entre señales de RMN, se pueden presentar tres casos de régimen de

intercambio (Figura 16):

1- Cuando el proceso de intercambio (asumido $A \rightleftharpoons B$) ocurre en un régimen más lento que la diferencia de los desplazamientos químicos entre las dos formas, cada núcleo afectado podrá asociarse a dos resonancias en el espectro en cuestión, correspondientes al mismo en dos entornos químicos diferentes.

2- Cuando el proceso de intercambio ocurre en un régimen rápido en la escala de tiempo de los desplazamientos químicos no se ve afectada la intensidad de la señal resultante pero sí su posición en el espectro, la cual resultará ser un promedio de los valores de desplazamiento químico esperados en cada entorno químico, promediado según la población relativa de cada uno.

3- En un régimen de intercambio químico intermedio en la escala de los desplazamientos químicos, se observará la coalescencia de ambas señales dando una señal promedio demasiado ancha como para ser detectada en el espectro de RMN; es decir, que tal intercambio se evidenciará por la ausencia de las resonancias esperadas o por un

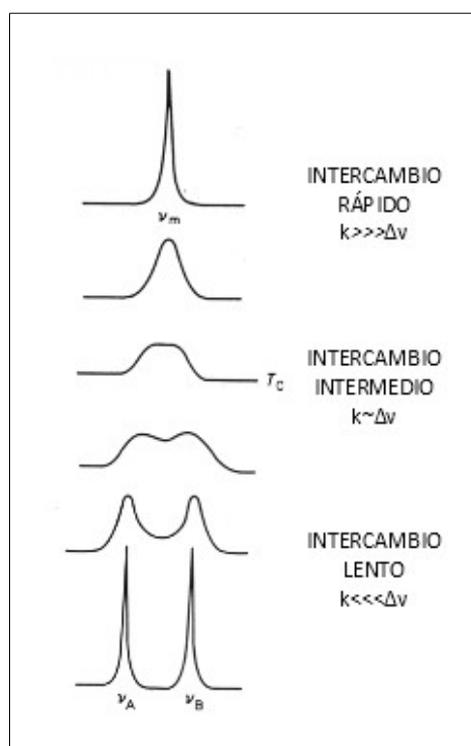


Figura 16: Representación esquemática de la evolución de una señal de RMN asociada a un núcleo susceptible a intercambio químico rápido, intermedio o lento en la escala de tiempo de los desplazamientos químicos entre los estados A y B. k es la constante de velocidad del intercambio y $\Delta\nu$ es la diferencia de frecuencias entre los estados A y B.

ensanchamiento de las mismas.

3.5.2 PROCEDIMIENTOS EXPERIMENTALES

Los espectros de resonancia magnética nuclear (RMN) se adquirieron a 298 K. Las muestras se suplementaron con agua deuterada (D_2O) 10% (Cambridge Isotope Laboratories). Los desplazamientos químicos secundarios se referenciaron con respecto a la señal del H_2O a 4,77 ppm (pH 6,8; 25°C) relativo a DSS, utilizando las relaciones de frecuencia 1H:X del punto cero de acuerdo a ¹⁸⁷.

3.5.2.1 ASIGNACIÓN DE RESONANCIAS DE DCL1-B

Los desplazamientos químicos de 1H , ^{13}C y ^{15}N del esqueleto peptídico de DCL1-B se asignaron utilizando los espectros de triple resonancia [BEST-HNCA/HN(CO)CA, BEST-HNCACB/HN(CO)CACB, BEST-HN(CA)CO/HNCO y HN(CA)HA] ¹⁸⁸ que se adquirieron en un espectrómetro Bruker Avance II 600, que opera a 600,13 MHz, empleando una sonda TXI de triple resonancia (1H , ^{13}C y ^{15}N) con detección inversa. La asignación de resonancias de los $H\alpha$ y $H\beta$, se realizó a partir de un espectro HBHA(CO)NH adquirido en un espectrómetro VARIAN 600 MHz equipado con una sonda criogénica.

Las resonancias de las cadenas laterales se asignaron a partir de espectros H(CC)(CO)NH y (H)CCH-TOCSY. La asignación se completó usando los espectros 1H - 1H -NOESY y 1H - 1H -TOCSY en 2H_2O . Las restricciones de distancia derivadas de espectros de tres dimensiones ^{15}N y ^{13}C -NOESY y 1H - 1H -NOESY (tm= 120 ms), y de dos dimensiones (2D) fueron adquiridos en un espectrómetro VARIAN VNMRS 800 MHz equipado con una criosonda de triple resonancia. Las resonancias correspondientes a los protones aromáticos se asignaron utilizando un espectro TOCSY 1H - 1H en 2H_2O . Para obtener la muestra en agua deuterada, se realizó una liofilización de la misma, y posteriormente se resuspendió en una solución amortiguadora preparada con 2H_2O . La asignación completa fue depositada en

BioMagResBank (entrada 18393).

3.5.2.2 DETERMINACIÓN DE RDCs DE DCL1-B

Para la adquisición de los parámetros de acoplamientos dipolares residuales (RDC), la muestra se alineó en una fase líquida cristalina formada por C12E5 y hexanol. Este medio preparado al 5% con una relación molar C12E5:hexanol de 0.96 crea una fase nemática en presencia del campo magnético del espectrómetro que induce un alineamiento preferencial de la proteína en la dirección del campo magnético¹⁸⁹. Los valores de RDC fueron medidos en tres pares de átomos de cada plano del péptido (D_{N-HN} , $D_{C'-HN}$, y $D_{C'-N}$) a partir de espectros ^{15}N -IPAP-HSQC acoplados en CO y desacoplados en $C\alpha$ en medio isotrópico y anisotrópico. En estos espectros para cada señal correspondiente a un par HN-N se obtienen cuatro picos que permiten medir simultáneamente las constantes de acoplamiento HN-N, N-CO(i-1) y HN-CO(i-1). El valor de RDC para cada residuo, se calculó a partir de la siguiente fórmula:

$$RDC = \Delta_{\text{isotrópico}} - \Delta_{\text{anisotrópico}}$$

donde $\Delta_{\text{isotrópico}}$ es igual a al acoplamiento en el medio isotrópico (J) y $\Delta_{\text{anisotrópico}}$ representa es el acoplamiento en el medio isotrópico más el acoplamiento dipolar residual (J+D) (Figura 17).

3.5.2.3 CÁLCULO DE ESTRUCTURA EN SOLUCIÓN POR RMN

La estructura de DCL1-B se calculó en base a las restricciones de distancia obtenidas de los espectros NOESY, y de los RDC. La señales de los espectros NOESY editados en ^{15}N y ^{13}C fueron seleccionadas automáticamente usando el algoritmo ATNOS-CANDID^{178,190} implementado en UNIO'10. La región aromática del espectro ^1H - ^1H -NOESY se seleccionó manualmente al igual que las señales del ^{15}N -NOESY que no se pudieron asignar automáticamente. Las constantes de acoplamiento escalar $^3J_{\text{HN-HA}}$ se determinaron a partir de un experimento HNHA. El

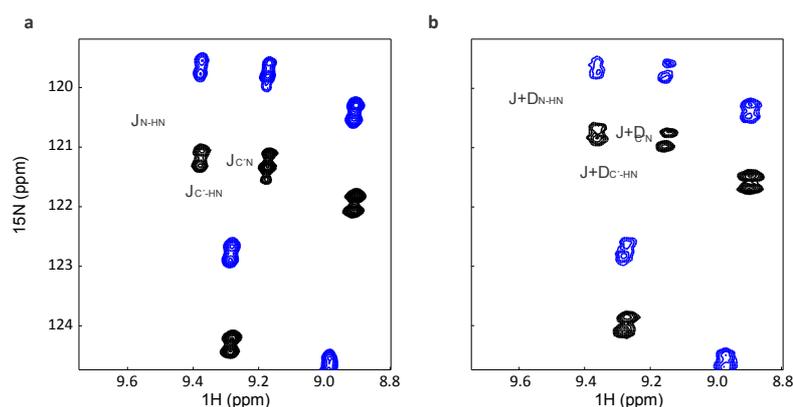


Figura 17: Zoom del espectro $^1\text{H}^{15}\text{N}$ -HSQC de la muestra proteica en medio a) isotrópico y b) anisotrópico usados para la determinación de RDC. Los componentes de desacoplamiento hacia arriba y hacia abajo se muestran en azul y negro, respectivamente.

análisis completo se llevó a cabo utilizando CCPNMR. Los ángulos de torsión ϕ y ψ se calcularon a partir de los valores de los desplazamientos químicos de H, N, $\text{C}\alpha$, $\text{C}\beta$, $\text{H}\alpha$ y CO usando la versión *online* de TALOS+¹⁹¹. y se utilizaron como restricciones adicionales. Las estructuras se calcularon con ARIA 2.3.1/CNS 1.1,^{192,193} permitiendo una reasignación de la lista de picos inicialmente cargada en UNIO'10. Los componentes axiales y rómbicos del tensor de alineamiento se determinaron con Pales¹⁹⁴ por ajuste de los RDC experimentales al conformero de menor energía del cálculo de estructura previo.

Se calcularon mil estructuras en el último ciclo iterativo de ARIA, de las cuales a los 20 conformeros de menor energía total se les realizó refinamiento con agua.

3.5.2.4 INTERACCIÓN PROTEÍNA-PROTEÍNA Y PROTEÍNA-ÁCIDOS NUCLEICOS

Los experimentos de interacción ARN/ADN-proteína y proteína-proteína se realizaron en un espectrómetro Bruker 600 MHz equipado con una consola

Avance II y una sonda TXI. Las titulaciones se realizaron en Hepes 50 mM (pH7), DTT 10 mM, Azida sódica 0,05 %, Complete 1 X y D₂O 10 %. El sustrato de ARN o ADN se calentó a 100°C y se enfrió inmediatamente en un baño de hielo y agua. Las titulaciones se realizaron agregando el ácido nucleico a una muestra de proteína 100 µM. En cada paso, se adquirió un espectro ¹H-¹⁵N-SOFAST-HMQC ¹⁹⁵.

La interacción de DCL1-B y DCL1-AB con HYL1-AB se evaluó por adición de 1 equivalente de HYL1-AB sin marca a una muestra de DCL1-B en PO₄³⁻ 100 mM, NaCl 50 mM y 1 mM β-mercaptoetanol (pH 7). Se adquirió un espectro ¹⁵N-HSQC para DCL1-B libre y otro para una mezcla 1:1 de DCL1-B y HYL1-AB. La concentración final de ambas proteínas fue de 110 µM.

3.5.2.5 DINÁMICA DE DCL1-B

Los parámetros de relajación de los ¹⁵N amídicos R1, R2 y NOEs se obtuvieron utilizando las secuencias de pulsos *hsqct13gpsi*, *hsqct2f3gpsi*, *hsqcnoe3f3gpsi*, respectivamente para obtener los espectros 2D.

Los tiempos de retardo o evolución usados en los espectros 2D fueron 10 (x2); 90; 200 (x2); 330; 450; 620; 830 y 1100 ms para los experimentos de T1 y, 16,16; 32,32 (x2); 48,48; 64,6; 80,8; 96,9; 113,12 (x2) y 129,29 ms para los experimentos T2. Los espectros ¹H ¹⁵N HSQC se adquirieron con en 2048 (¹H) y 128 (¹⁵N) puntos complejos, el ¹H ¹⁵N NOE se adquirieron con 2048 (¹H) y 256 (¹⁵N) puntos complejos.

Una vez obtenidos los parámetros T1, T2 y NOE ¹H-¹⁵N, se calcularon los parámetros de movilidad de la proteína utilizando el programa TENSOR2 ¹⁸³ y realizando un análisis de "model free".

3.6 SÍNTESIS DE ARN

Las muestras de ARN se sintetizaron por transcripción *in vitro* por la ARN

polimerasa T7, usando oligonucleótidos hibridados. La mezcla se preparó con un tampón de transcripción 1X [Tris 40 mM (pH 8), DTT 5 mM, espermidina 1 mM, Tritón X-100 0,01 % y PEG 8000 80 mg/ml], cada rNTP a una concentración de 4 mM (rA, rC, rG y rU), MgCl₂ 20 mM, BSA 40 µg/ml y 1 U de pirofosfatasa. La reacción se inició con el agregado de la enzima T7 polimerasa y continuó por 3 horas a 37°C. Luego, se agregaron 50 U de ADNasa y la mezcla se incubó por 30 minutos a 37°C. La mezcla de reacción se diluyó 8 veces en Tris 20 mM (pH 8), EDTA 10 mM y urea 8M y se cargó en una columna "Q-Sepharose" equilibrada con la solución amortiguadora mencionada previamente. La columna se eluyó con un gradiente de 0 a 1 M NaCl en el tampón. Las fracciones que contenían ARN se determinaron por A₂₆₀, y se chequearon por electroforesis en gel de poliacrilamida 5 % con Urea. Las fracciones que contenían el transcripto deseado se dializaron contra 200 volúmenes de H₂O.

3.7 ESPECTRO DE EMISIÓN DE FLUORESCENCIA

El espectro de emisión de fluorescencia de DCL1-B se realizó sobre una muestra de concentración 5 µM contenida en una solución amortiguadora de PO₄³⁻ 100 mM (pH7) y NaCl 50 mM. Para la adquisición de los espectros, se utilizó un espectrofluorómetro Varian Cary Eclipse fijando previamente los valores de λ máxima de emisión excitando a 280 nm y de λ máxima de excitación. Este último, arrojó un valor de 282 nm y fue utilizado para obtener el espectro de emisión de fluorescencia en un rango de 300 a 500 nm.

3.8 TITULACIONES POR ANISOTROPÍA DE LA FLUORESCENCIA

Para las titulaciones por anisotropía de fluorescencia, los fragmentos de ARN se marcaron con fluoresceína usando el sistema *EndTag Nucleic Acid End Labeling* y la marca reactiva fluoresceína maleimida de Vector Laboratories. Los fragmentos marcados se purificaron con una extracción con fenol, se precipitaron

con etanol y se resuspendieron con una solución amortiguadora de PO_4^{3-} 10 mM pH7. El ADN marcado con fluoresceína se obtuvo de Sigma. La anisotropía de fluorescencia se midió en un espectrofluorómetro Varian Cary Eclipse excitando las muestras a 492 nm y midiendo la emisión de fluorescencia a 520 nm. Los valores de anisotropía se obtuvieron por un promedio de 3 medidas con un tiempo de integración de 20 segundos. La apertura de excitación y emisión usadas fueron de 10 nm. El ARN fue calentado a 100°C por 5 minutos y enfriado inmediatamente en un baño con hielo y agua para asegurar la hibridación molecular correcta. Los oligómeros de ADN se hibridaron mezclando ambas hebras y calentando a 100°C dejando enfriar lentamente a T° ambiente. La solución amortiguadora de cada muestra se cambió por PO_4^{3-} 10 mM pH7 inmediatamente antes de realizar las mediciones. El cambio se realizó utilizando columnas Bio-Spin 6 (Bio-Rad) de acuerdo a las indicaciones del proveedor. Las curvas de unión se realizaron a 20°C titulando una solución de ácido nucleico 50 nM marcado, con cantidades crecientes de proteína. Los datos experimentales (r) se ajustaron a una función hiperbólica (ecuación 1) asumiendo un modelo de unión simple con una estequiometría 1:1 o a la ecuación de Hill (ecuación 2):

$$r = r_0 + \frac{a[P]}{b + [P]} \quad [1]$$

$$r = r_0 + \frac{a[P]^n}{b^n + [P]^n} \quad [2]$$

donde [P] corresponde a la concentración de proteína libre, r_0 es la anisotropía del ARN o ADN libre, a es la amplitud del cambio de anisotropía tras la unión, b es la constante de disociación y n es el coeficiente de Hill que da una medida de la cooperatividad de unión para sistemas con estequiometría mayor que 1:1. Las curvas de titulación se normalizaron restando el valor de r_0 y dividiendo por la amplitud a de cada punto .

3.9 DICROÍSMO CIRCULAR

La elipticidad de las muestras de proteínas se evaluó utilizando un espectropolarímetro Jasco 810 calibrado con (+) ácido 10-canforsulfónico. Para el espectro de dicroísmo circular (CD) en el UV cercano se utilizó una concentración de proteína de 80 μM y el espectro se realizó en un rango de 250 a 350 nm, mientras que para la adquisición del espectro en el UV lejano, la concentración de la muestra proteica utilizada fue de 10 μM y se adquirieron lecturas en un rango entre 190 y 250 nm. Ambas muestras proteicas se encontraban disueltas en una solución amortiguadora de PO_4^{3-} 100 mM (pH7) y NaCl 50 mM.

Para los experimentos de estabilidad por CD, las proteínas replegadas se desalaron en columnas de desalado PD-10 (GE Healthcare) usando una solución amortiguadora de PO_4^{3-} 10 mM (pH7). Las concentraciones de proteínas se determinaron por espectroscopía UV. Se prepararon soluciones de proteína a distintas concentraciones de urea y se incubaron a T^o ambiente durante tres horas para asegurar condiciones de equilibrio y minimizar las modificaciones químicas. Los espectros de CD en el UV lejano se registraron en un rango entre 190 y 250 nm, el desplegamiento inducido por urea se monitoreó por cambios de la elipticidad a 220 nm. La concentración de proteínas fue de 10 μM , y el paso óptico de la celda utilizada fue de 0,1 cm. En todos los casos los puntos se adquirieron a una velocidad de 20 nm min⁻¹ y se promediaron al menos 3 medidas para cada muestra. La señal a 250 nm se utilizó como control interno para corregir pequeñas fluctuaciones en la línea de base. Todas las medidas se realizaron a 20°C.

3.10 OBTENCIÓN DE PLANTAS DE *A. thaliana* TRANSFORMADAS

3.10.1 ECOTIPO UTILIZADO Y CONDICIONES DE CRECIMIENTO

Las plantas de *A. thaliana* utilizadas en esta tesis corresponden al ecotipo Columbia-0 (Col-0). Las plantas fueron cultivadas en una cámara de crecimiento con un régimen de 16 horas de luz ($100 \mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) y 8 horas de oscuridad. La temperatura de crecimiento fue de 23°C durante el ciclo luz/oscuridad, mientras que la humedad fue mantenida en 65 % de humedad relativa. Las plantas fueron regadas 2 veces por semana con agua. Para el crecimiento directo en tierra, las semillas fueron estratificadas a 4°C por 2 días en tubos de microcentrífuga con 1 ml de agar 0,1% (p/v), y luego sembradas en tierra. Las semillas de las líneas mutantes *hyl1-2* (SALK_064863), fueron obtenidas a través del “*Arabidopsis Biological Resource Center*” (ABRC).

3.10.2 TRANSFORMACIÓN DE *A.thaliana*

La transformación de plantas de *A. thaliana* fue realizada mediante la técnica de “*floral-dip*”¹⁹⁶. Se cultivaron las cepas de *A. tumefaciens* ASE transformadas con los distintos plásmidos binarios en 5 ml de LB conteniendo 80 $\mu\text{g}/\text{ml}$ de espectinomicina, 50 μg de kanamicina y 25 μg de cloramfenicol a 28°C hasta llegar a una DO_{600} cercana a 2. Se concentraron las células por centrifugación a 5.000 *g* durante 10 min a temperatura ambiente y se resuspendieron en un medio con sacarosa 5 % (p/v) y 200 $\mu\text{l}/\text{l}$ de Silwet L-77. Las inflorescencias de plantas de *A. thaliana* regadas el día anterior fueron sumergidas en la suspensión de células durante 15 segundos. Luego de escurrir el exceso de medio de cultivo, se colocaron las plantas en posición horizontal dentro de bandejas cubiertas con film para conservar la humedad. Luego de 48 horas se quitó el film, se colocaron en posición vertical y se prosiguió con el cultivo en la cámara de crecimiento hasta la maduración de las semillas.

3.10.3 RECOLECCIÓN DE SEMILLAS

La recolección de semillas se realizó a partir de plantas que finalizaron su ciclo de vida y que presentaban semillas maduras. Se retiraron las plantas del pote de tierra y frotando las mismas con las manos se dejó caer las semillas sobre una hoja de papel. Se eliminaron los restos de silicua y hojas por tamizado. Las semillas se almacenaron en micotubos de 1,5 ml hasta su posterior utilización.

3.10.4 ESTERILIZACIÓN DE SEMILLAS

Para la esterilización de las semillas, las mismas fueron incubadas con una solución de etanol 70 % (v/v) x, tritónX-100 0,1 % (v/v), durante 15 minutos. Luego se realizó un enjuague con etanol 96 % (v/v) y finalmente se las dejó secar en el flujo laminar. La manipulación del material vegetal en condiciones de esterilidad se realizó en una cabina de flujo laminar.

3.10.5 SELECCIÓN DE TRANSFORMANTES

Para la selección de plantas transformadas con el vector binario pCHF3-CS052, semillas esterilizadas fueron sembradas en placas de Petri con medio MS-agar¹⁹⁷ suplementado con 50 µg/ml de kanamicina. Las placas fueron mantenidas en oscuridad a 4°C por 48 horas. Transcurrido este periodo las placas fueron transferidas a la cámara de cultivo y en estas condiciones las plantas se dejaron crecer 10 días. Luego las placas fueron analizadas y aquellas plantas que presentaban un fenotipo resistente al agente de selección, evidenciado por el desarrollo de una raíz de tamaño considerable y del primer par de hojas verdaderas, fueron transferidas a potes con tierra. Los potes fueron ubicados en bandejas con agua las cuales se cubrieron con film de nylon transparente, de modo de evitar la pérdida de humedad. En estas condiciones se cultivaron en la cámara durante 2 días, momento en el que se retiró el nylon y se continuó con su cultivo. Aquellas plantas transformadas con el vector binario pCHF5 fueron

incubadas en tubos de microcentrífuga en una solución de agar 0,1% (p/v) a 4°C en oscuridad durante 48h. Luego fueron sembradas directamente en potes con tierra. Los potes fueron ubicados en bandejas con agua las cuales se cubrieron con film de de nylon transparente, de modo de evitar la pérdida de humedad. En estas condiciones se cultivaron en la cámara durante 2 días, momento en el que se retiró el nylon y se continuó con su cultivo. A partir de ese momento fueron regadas 2 veces por semana con una solución acuosa de herbicida glufosinato de amonio 200 mg/l.

3.11 MUESTRAS VEGETALES

3.11.1 RECOLECCIÓN DE MUESTRAS

Se extrajeron inflorescencias cerradas de plantas de aproximadamente 2 meses de vida y manteniendo los tubos de almacenamiento sumergidos en N₂ líquido. Una vez recolectadas, las muestras se almacenaron a -70 °C hasta su utilización.

3.11.2 EXTRACCIÓN DE PROTEÍNAS DE INFLORESCENCIAS

Las inflorescencias se homogeneizaron sumergiendo microtubos de 1,5 ml en N₂ líquido y rompiendo el tejido con un pilón. A continuación se adicionaron 5 µl de solución amortiguadora cada 1 mg de tejido. El tampón de extracción está compuesto por Tris-Hcl 50 mM (pH 7,8), MgCl₂ 1 mM, EDTA 2 mM, PMSF 1 mM, Ascorbato 2 mg/ml y Glutación reducido 0,3 mg/ml. Los homogenados se centrifugaron a máxima velocidad por 30 min a 4°C. El sobrenadante se trasvasó a un tubo limpio.

3.11.3 CUANTIFICACIÓN DE PROTEÍNAS

La concentración de proteínas de los extractos vegetales se determinó por el método de Bradford ¹⁹⁸. El ensayo consistió en agregar 0,5 ml de solución de

azul de Coomasie G-250 a 0,5 ml de solución proteica, mezclar rápidamente y determinar la absorbancia a 595 nm en un espectrofotómetro. Como testigo se utilizó albúmina sérica bovina de concentración conocida en el rango de 0 a 8 µg/ml.

3.12 WESTERN BLOT

Las proteínas separadas por geles de poliacrilamida en condiciones desnaturizantes se electrotransfirieron a membranas de nitrocelulosa (Amersham) utilizando el sistema Mini-Trans-Blot transfer (Bio-Rad). La transferencia se realizó con agitación permanente en baño de agua y hielo durante 1 hora a 350 mA o, alternativamente, durante toda la noche a 14 V y 4°C. La solución de transferencia empleada consiste en Tris-HCl pH 8 25mM, glicina 192 mM y metanol 20% (v/v). Concluida la transferencia, se bloqueó la membrana incubandola con una solución amortiguadora PBS-T leche (NaCl 137 mM, Fosfato 10 mM, KCl 2.7 mM, Tween 20 0.1 % v/v, leche descremada en polvo de origen comercial 5% p/v) durante 1 hora con agitación suave a temperatura ambiente (o toda la noche a 4°C). Posteriormente la membrana se incubó toda la noche a 4°C con agitación suave (o 1 hora a temperatura ambiente) con el anticuerpo primario comercial anti-HA de conejo (Sigma). Posteriormente, se realizaron 4 lavados de 10 minutos con PBS-T con agitación suave para eliminar los anticuerpos no unidos a la membrana y se incubó la misma con el anticuerpo secundario durante 1 hora a temperatura ambiente (o toda la noche a 4°C) con agitación suave. Luego, se efectuaron 4 lavados de 10 minutos con una solución de lavado que dependió del sistema utilizado para el revelado. En los casos en los que se indica, se utilizó el revelado por fosfatasa alcalina. En este, la membrana se lavó con solución para revelado por fosfatasa alcalina (Tris-HCl 100 mM (pH 9.5), NaCl 100 mM, MgCl₂ 5 mM), y luego se incubó con 10 ml de dicha solución conteniendo 40 µl de NBT 50 mg/µl y 20 µl de

BCIP 20 mg/ μ l, a oscuras, durante el tiempo necesario hasta que se evidenció la señal, generalmente entre 5 y 10 minutos. La reacción se detuvo por sucesivos lavados de la membrana con de H₂O destilada.

Cuando fue requerida una mayor sensibilidad en la detección de la señal, se reveló empleando un sistema quimioluminiscente. Para esto se utilizó el sistema ECL Western Blotting detection System de Amersham® siguiendo las instrucciones del fabricante.

3.13 MICROSCOPIA CONFOCAL.

Para obtener las imágenes de las raíces por microscopía confocal, se utilizaron plántulas crecidas en medio MS. Luego de la germinación, las plántulas fueron crecidas colocando las placas de Petri en posición vertical con el objetivo de favorecer el crecimiento radicular y de causar el menor daño posible al utilizarlas. El tiempo de cultivo previo a la obtención de las imágenes fue de siete días.

Las raíces de las plántulas fueron incubadas por 1 minuto en Ioduro de Propidio (IP) 1 mg/ml. En estas condiciones, el IP se une a las paredes celulares, emitiendo fluorescencia roja. A continuación, fueron montadas en agua entre porta- y cubreobjetos. Finalmente, las imágenes al microscopio confocal se obtienen utilizando un objetivo 20X en un microscopio Nikon Eclipse TE-2000-E2. Se utilizó un laser multilinea de ión de argón (488nm) para la excitación, y filtros de 515/30 para la fluorescencia de GFP y de 605/75 para la fluorescencia de IP. El procesamiento de imágenes se realizó con el programa EZ-C1 Gold Version de Nikon Corporation.

3.14 PROGRAMAS UTILIZADOS

Todos los espectros de RMN fueron adquiridos el programa TopSpin de

Bruker en su versión 2.1 y procesados con el programa *NMRPipe*¹⁹⁹. El análisis de los espectros de asignación del esqueleto peptídico se realizó con *CARA*²⁰⁰ y de las cadenas laterales con *NMRView*²⁰¹. La asignación de los NOE se realizó utilizando el programa UNIO'10, versión 2.0.1, que combina los programas ATNOS/CANDID (versión 2.0)¹⁷⁸. El cálculo de las estructuras a partir de los valores de RDC para DCL1-B se realizó mediante el programa CS-Rosetta^{202,203}, mientras que el modelado de las estructuras de las mutantes de HYL1-A en función de la secuencia se llevó a cabo con el programa Rosetta^{204,205}.

El análisis de los espectros de $^1\text{H}^{15}\text{N}$ -HSQC utilizados para la determinación de la interacción proteína-ligando y los espectros utilizados para la determinación de los parámetros de relajación se realizó con CCPNMR²⁰⁶. Los datos de relajación (R1, R2, NOE heteronuclear) fueron analizados según el formalismo *Model Free*²⁰⁷ empleando el programa Tensor v2.0¹⁷⁹. El análisis de las estructuras tridimensionales de las proteínas estudiadas se realizó con el programa PyMOL versión 0.96²⁰⁸. La determinación del potencial electrostático mapeado sobre las estructuras se realizó con la herramienta APBS²⁰⁹ implementada en pymol. Las correlaciones entre los RDC observados y medidos se realizaron con el programa MODULE 1.0²¹⁰.

4 CAPÍTULO 1: dsRBD2 de DCL1

4.1 RESULTADOS 1

4.1.1 RESULTADOS PREVIOS

DCL1 es una enzima que pertenece a la familia de ARNasas III y junto con las proteínas accesorias HYL1 y SERRATE, lleva a cabo el procesamiento de precursores de miARNs de manera coordinada y precisa en *A. thaliana*. Dentro del complejo de procesamiento, DCL1 es responsable de realizar los cortes en los precursores que dan lugar a los miARN, mientras que HYL1 y SE, son proteínas auxiliares cuya función principal es aumentar la precisión de DICER. Debido a su importancia durante el proceso de biogénesis de miARNs, la caracterización de DCL1 resulta de gran interés. Sin embargo, esta proteína tiene un tamaño de 1909 aminoácidos y una arquitectura de dominios compleja, lo que dificulta en gran medida los estudios estructurales sobre DCL1 completa. Por este motivo, la mayoría de los análisis estructurales se realizan utilizando fragmentos de la proteína definidos por los motivos que contiene. DCL1, posee dos dominios de unión a ARNdh en el extremo C-terminal que podrían conferirle la capacidad de unir los precursores de miARN. Con el objetivo de caracterizar estos dominios, se obtuvieron en el laboratorio varias construcciones que corresponden a los motivos individuales y a combinaciones de los mismos (Figura 18).

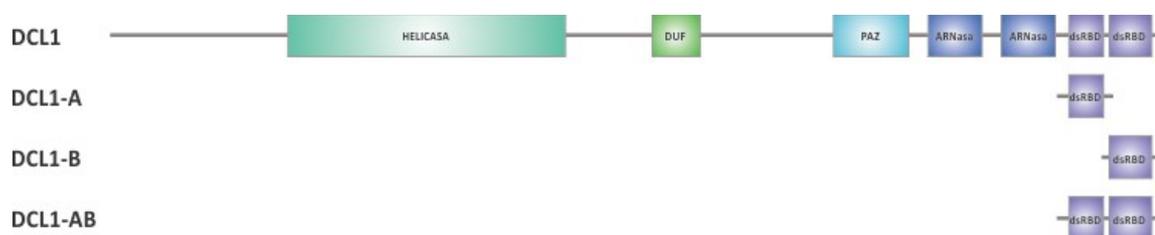


Figura 18: Construcciones de los dominios dsRBDs de DCL1 de *A.thaliana*

Luego de expresar cada proteína y realizar un espectro de RMN de correlación $^1\text{H}^{15}\text{N}$, se pudo determinar que el segundo dsRBD de DCL1 (DCL1-B) da espectros $^1\text{H}^{15}\text{N}$ -HSQC bien resueltos, mientras que el primero, DCL1-A, se corresponde con el de una proteína desestructurada, lo que se evidencia por la baja dispersión de las señales en la dimensión de ^1H en el espectro. La expresión de los dos dominios juntos no ayudaría al plegamiento del primer dsRBD, debido a que todas las señales del doble dominio se encuentran perfectamente solapadas con las de los dominios individuales (Figura 19).

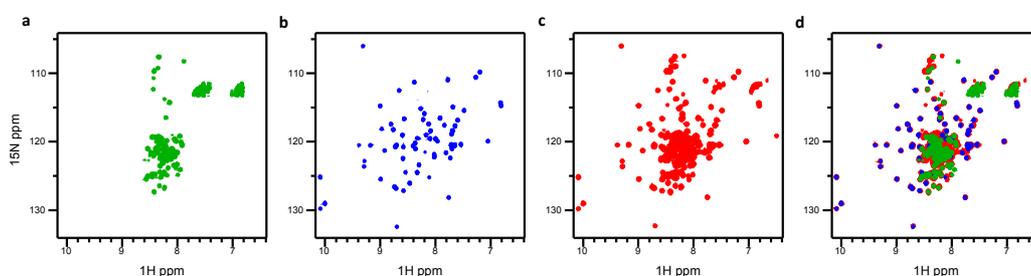


Figura 19: Espectros de correlación de RMN de $^1\text{H}^{15}\text{N}$ de las construcciones a) DCL1-A; b) DCL1-B; c) DCL1-AB y d) superposición de los espectros a, b y c.

A partir de estos resultados, nos propusimos enfocar nuestros estudios en DCL1-B. Como se mencionó en la sección 1.2.2.1.1, la proteína DCL1 de plantas se caracteriza por tener un dsRBD extra en el extremo C-terminal con respecto a las proteínas DICER de animales. Este segundo dsRBD se encuentra altamente conservado entre las distintas especies de plantas, lo que lo convierte en un buen candidato para llevar a cabo funciones de reconocimiento específico de sustratos. Una característica que diferencia a DCL1-B de otros dsRBDs canónicos es la presencia de una inserción de seis aminoácidos en el bucle $\beta 2$ - $\beta 3$ (Figura 20) que posiblemente pueda tener alguna razón funcional.

La construcción de DCL1-B generada en el laboratorio tiene 85 aminoácidos, un peso molecular de 9,554 KDa y un punto isoeléctrico de 9,45. En la figura 21 se muestra la secuencia de aminoácidos de la misma.

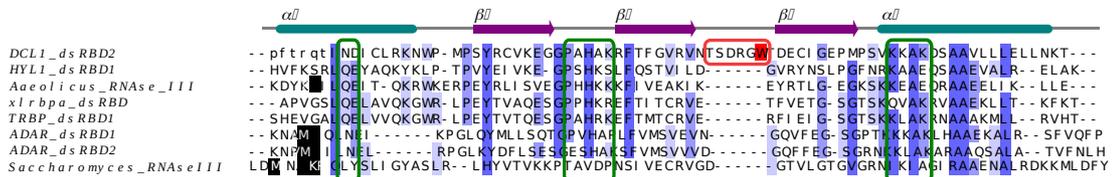


Figura 20: Alineamiento de la secuencia de DCL1-B con secuencias de dsRBDs para los cuales está disponible la estructura en complejo con el ARN. Los elementos de estructura secundaria se muestran en la parte superior del alineamiento. Los residuos resaltados en azul son los que presentan cierto grado de conservación. Los recuadros verdes resaltan los tres motivos de interacción con el sustrato. Los residuos de la hélice $\alpha 1$ que participan en la unión con el ARNdh están resaltados con recuadros negros. El cuadro rojo remarca la inserción de DCL1-B y el Trp 55 se muestra con fondo rojo.

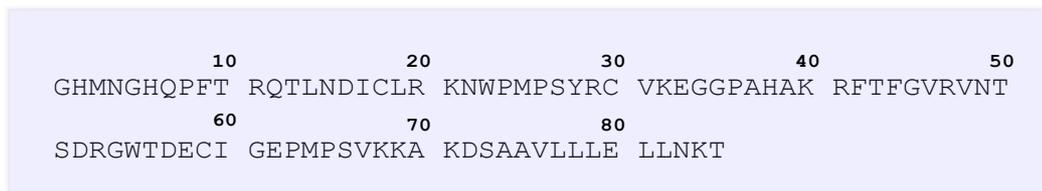


Figura 21: Secuencia de aminoácidos de la construcción de DCL1-B

4.1.2 OPTIMIZACIÓN DE CONDICIONES PARA EL ESTUDIO DE LA PROTEÍNA POR RMN

Con el objetivo de obtener muestras de una concentración adecuada para realizar experimentos de RMN se optimizaron las condiciones de los protocolos de expresión y purificación de la proteína marcada con ^{15}N y $^{15}\text{N}^{13}\text{C}$.

4.1.2.1 EVALUACIÓN DE PROTOCOLOS DE EXPRESIÓN

Para evaluar la expresión de la proteína DCL1-B, se llevaron a cabo dos procedimientos diferentes; el método de inducción tradicional por IPTG y el método de inducción con IPTG de alta densidad ²¹¹. Los mismos se muestran esquematizados en la figura 22.

Al evaluar ambos protocolos de expresión, se pudo observar que en el caso de la inducción tradicional, la OD₆₀₀ alcanzada transcurridas las 4 horas desde la adición de IPTG es de aproximadamente 1,5-2,5 en medio M9, mientras que para el caso del protocolo de alta densidad, la OD₆₀₀ final es de 10-15 en el mismo medio de cultivo. A pesar de estas diferencias en la densidad óptica final, los rendimientos de proteína recombinante fueron similares en ambos casos. Al analizar los espectros ¹H de la proteína recombinante

expresada a partir del protocolo de alta densidad, se notó la presencia de dobletes y singletes en los picos correspondientes a protones amida del esqueleto peptídico. Este efecto, es indicativo de un bajo rendimiento en la incorporación de la marca isotópica en la proteína. Luego de repetir el método

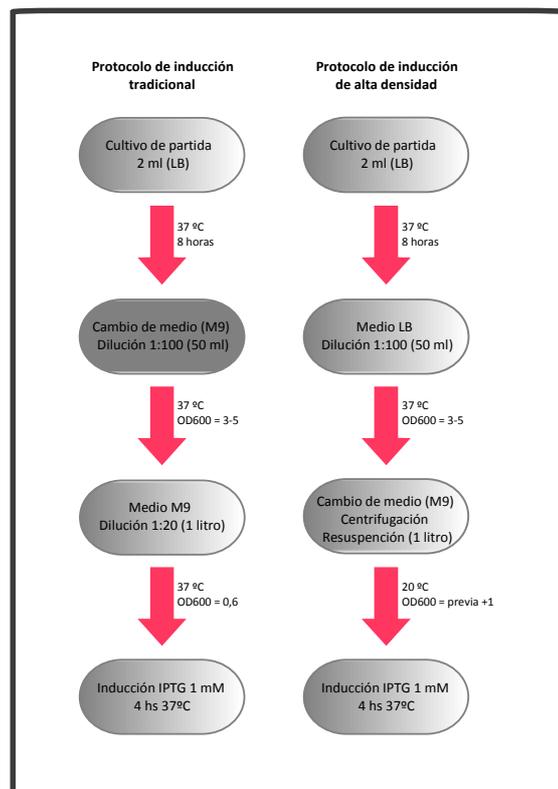


Figura 22: Método de expresión tradicional por inducción con IPTG y método de inducción por IPTG de alta densidad.

de expresión de alta densidad en varias oportunidades se observó que el porcentaje de marca que incorpora la proteína recombinante tras cada expresión es aleatoria. Por este motivo, se utilizó el protocolo de inducción tradicional en las sucesivas purificaciones de DCL1-B y de DCL1-AB.

4.1.2.2 EVALUACIÓN DE PROTOCOLOS DE PURIFICACIÓN

Uno de los problemas que se presentó durante la purificación de la proteína fue la elevada cantidad de ácidos nucleicos presentes en la muestra. La presencia de los mismos impidió la obtención de señales de la proteína durante la adquisición de los espectros de RMN debido a la formación de complejos inespecíficos muy grandes. Con el objetivo de eliminar los ácidos nucleicos de las muestras, por un lado, se llevó a cabo un protocolo de purificación de proteínas desnaturalizadas como se describe en la sección 3.4.3, y por otro se evaluó el efecto de distintos tipos de cromatografías. Al realizar la **purificación desnaturalizante**, definitivamente se obtuvieron muestras libres de ácidos nucleicos, sin embargo, el rendimiento de proteína fue muy bajo, sufriendo la mayor pérdida durante la etapa de replegado. En el caso de las **cromatografías**, se evaluaron tres tipos distintos: de exclusión molecular (columna Superdex long 10/30), de intercambio catiónico (columna Hiprep 16/10 CM) y de intercambio aniónico (columna MonoQ). En todos los casos las columnas fueron equilibradas con Tris-HCl 50 mM (pH 8), NaCl 50 mM y β -mercaptoetanol 1 mM. Luego de realizar los cromatogramas, se observó que todas las muestras se obtuvieron libres de ácidos nucleicos. En el caso de la cromatografía de exclusión molecular, se realizó una calibración con distintas proteínas de peso molecular conocido, y se pudo observar que DCL1-B da una banda de elución a un volumen esperado de acuerdo con su peso molecular y que además, la proteína se obtiene en estado monomérico, lo que se evidencia por la presencia de un solo pico correspondiente a la fracción proteica. Al comparar los distintos cromatogramas,

se evidenció que los mayores rendimientos se alcanzaron utilizando la columna de intercambio catiónico (Hiprep 16/10 CM) (Figura 23).

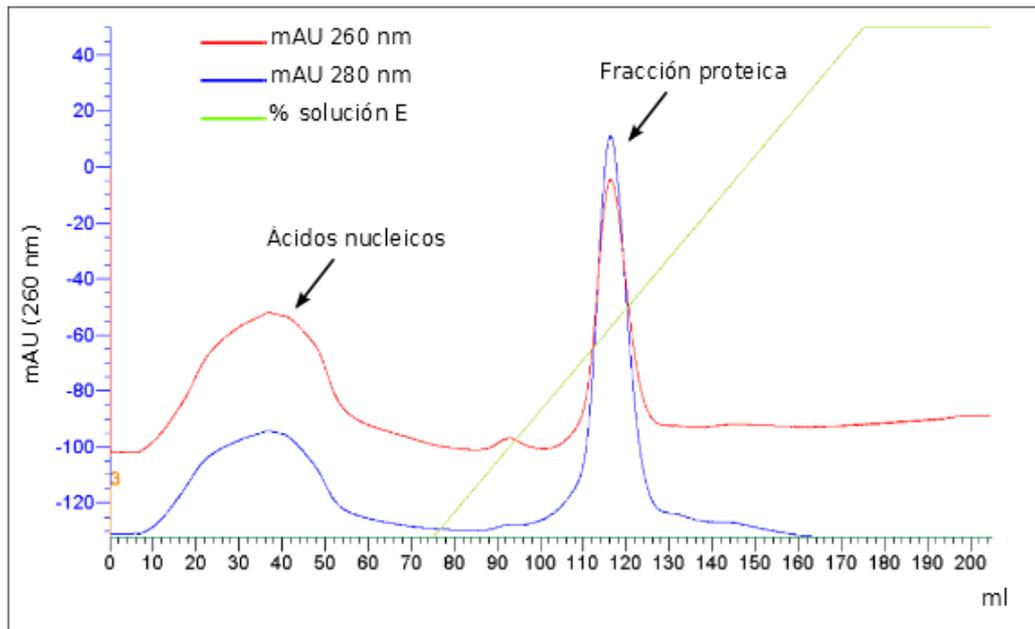


Figura 23: Cromatografía de intercambio iónico utilizando una columna Hiprep 16/10 CM. En azul se indica la absorbancia a 280 nm, en rojo, la absorbancia a 260 nm y en verde el gradiente de NaCl (de cero a 100%) utilizado. Los picos correspondientes a las fracciones de ácidos nucleicos y proteicas están indicados con flechas.

Por otro lado, se notó que durante las distintas etapas de la purificación la proteína precipitaba en grandes cantidades, lo que llevaba a una disminución significativa de la concentración final. Por este motivo, se decidió evaluar la influencia del pH sobre la muestra. Los diferentes valores de pH utilizados fueron 5, 6 y 7. La mayor estabilidad de la proteína se logró a pH 7, donde no se observó precipitación. También se evaluó la influencia de la digestión de la etiqueta de histidina de la proteína de fusión. Al evaluar el nivel de precipitado de DCL1-B sin digerir y digerida, se observó que las proteínas digeridas presentaban mayor solubilidad.

A partir de estos resultados, se obtuvo un protocolo de purificación optimizado respecto a rendimiento y calidad de la muestra. De esta manera, se logró obtener una cantidad suficiente de proteína (ca. 1,5 mg de proteína en un litro de cultivo en medio M9) con altos niveles de pureza (Figura 24). Las muestras obtenidas a partir de este protocolo fueron concentradas y se les agregó azida 0,05%, 1 mM TCEP y 10% D₂O para utilizarlas en la adquisición de espectros de RMN.

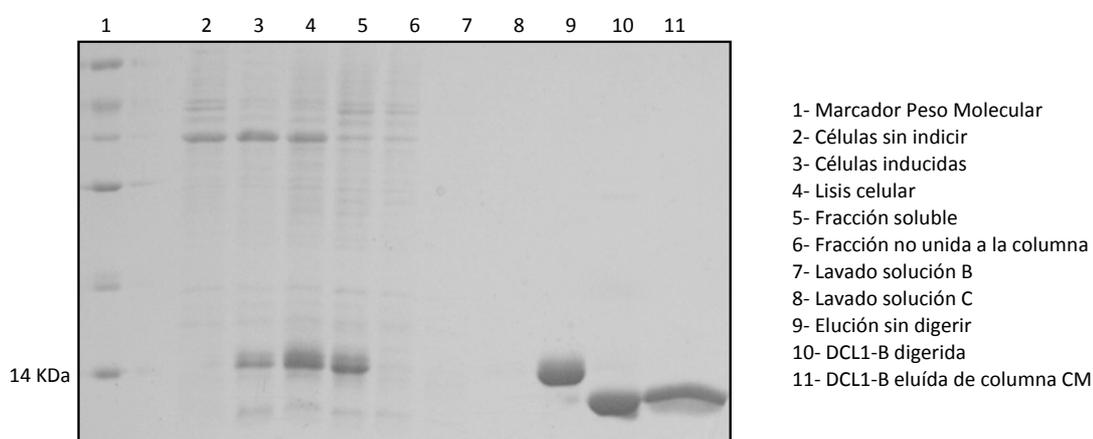


Figura 24: Gel de poliacrilamida 15% en condiciones desnaturizantes para el seguimiento de la purificación de la proteína doble marca. A la derecha se indica lo sembrado en cada calle. A la izquierda se indica el tamaño correspondiente a la banda de marcador de peso molecular. La composición de las soluciones B y C se encuentra en la sección 3.4.3 de materiales y métodos.

4.1.3 ESPECTROS DE FLUORESCENCIA Y DICROÍSMO CIRCULAR

DCL1-B tiene en su secuencia cuatro residuos de fenilalanina, dos triptofanos y una tirosina. La espectroscopía electrónica de los residuos aromáticos de las proteínas contiene información estructural útil, por lo que se realizó una caracterización espectroscópica de este dominio utilizando las técnicas de fluorescencia y dicroísmo circular. En primer lugar, se realizó un

espectro de emisión de fluorescencia utilizando una λ de excitación de 282 nm (Figura 25). El espectro muestra una banda de emisión a 345 nm lo que indica que al menos uno de los dos residuos de triptofano está expuesto al solvente. Con el objetivo de obtener información estructural adicional de la proteína, se realizaron espectros de dicroísmo circular (CD). En el espectro adquirido en el UV lejano, se observó la estructura de bandas característica de proteínas con estructura secundaria con estructura secundaria $\alpha+\beta$, correspondiente a lo esperado para un dsRBD. En el espectro de CD en el UV cercano, se observó la presencia de una banda relativamente intensa con estructura fina en la región de 250-300 nm que indica que al menos alguno de los residuos aromáticos de DCL1-B están localizados en regiones estructuradas de la proteína, en un ambiente rígido y asimétrico (Figura 25).

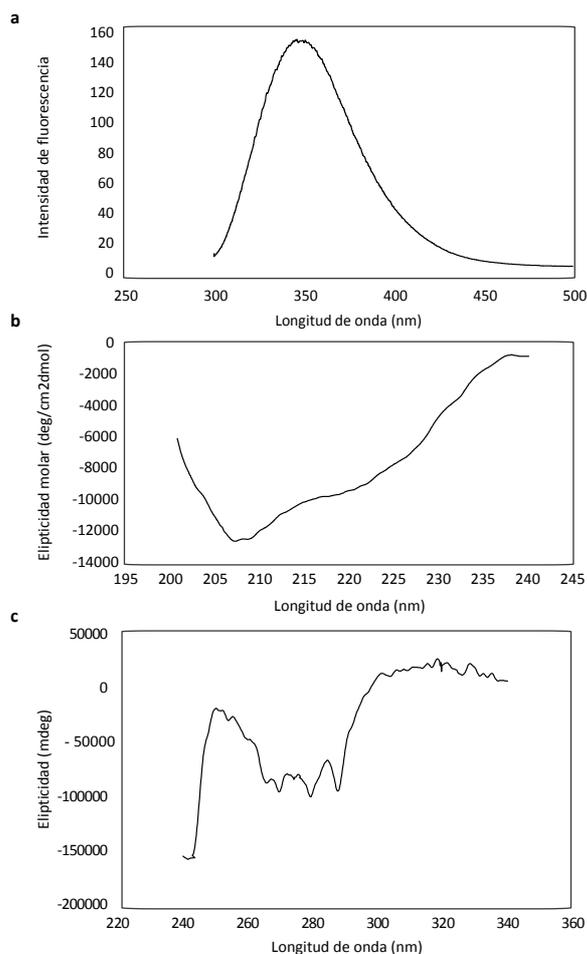


Figura 25: Figura 8: a) Espectro de emisión de fluorescencia de DCL1-B excitando a 346 nm. b) Espectro de dicroísmo circular en el UV lejano que muestra la estructura secundaria de tipo $\alpha+\beta$. c) Espectro de dicroísmo circular en el UV cercano. Se observan las bandas características de aromáticos entre 250-300 nm.

4.1.4 ASIGNACIÓN DE LAS RESONANCIAS DE DCL1-B Y CSI

Tras obtener una muestra adecuada para realizar estudios por RMN, se llevó a cabo la asignación del espectro $^1\text{H}^{15}\text{N}$ -HMQC mediante la adquisición de experimentos de triple resonancia. Luego de la adquisición de los espectros, la muestra mostró cierto grado de degradación, lo que pudo evidenciarse con el aumento del número de señales en la región central de los $^1\text{H}^{15}\text{N}$ -HMQC adquiridos luego de cada uno de los experimentos. No todas las señales del espectro $^1\text{H}^{15}\text{N}$ -HMQC, pudieron ser asignadas a grupos amida del esqueleto proteico. En la figura Figura 36, se pueden observar las señales correspondientes

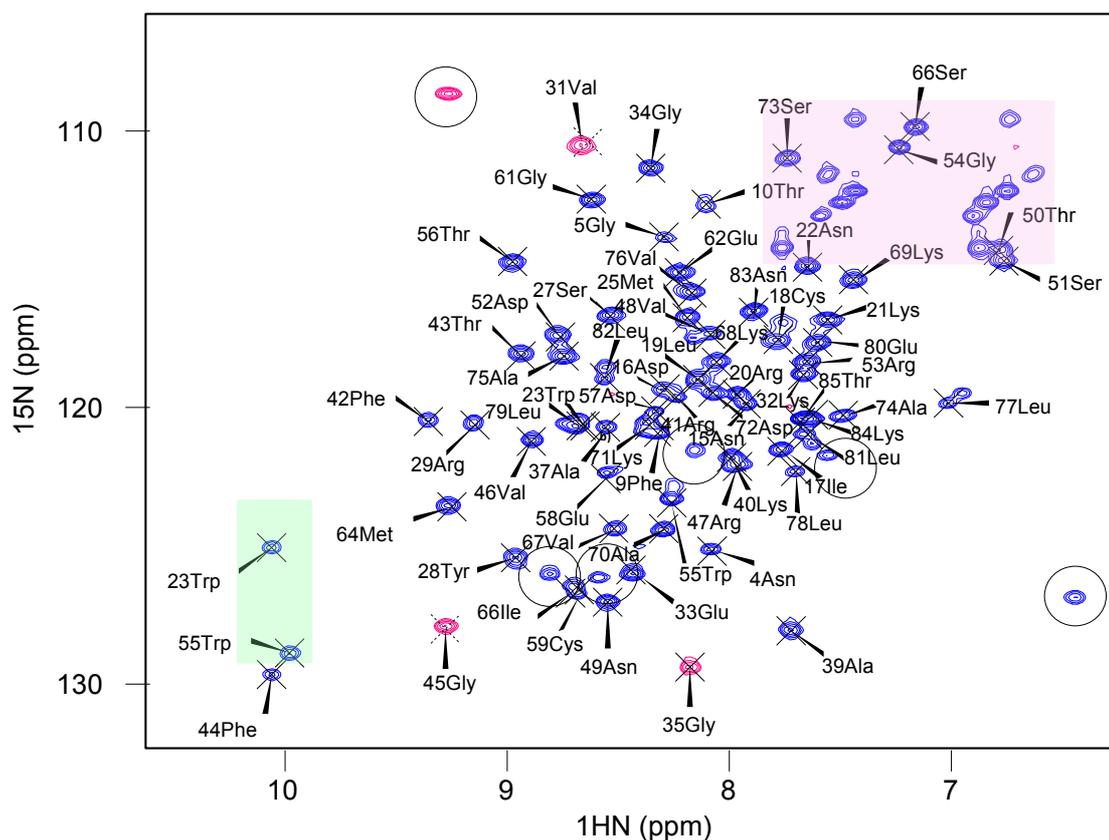


Figura 26: Espectro SOFAST ^{15}N -HMQC de DCL1-B asignado. Las señales azules corresponden a señales positivas y en rosa las negativas. Recuadradas en rosa se encuentran las señales correspondientes a cadenas laterales de Asn y Gln, en verde, señales del grupo indol de los dos Trp y resaltadas en círculos las señales que no se pudieron asignar. La muestra utilizada tenía una concentración de 600 μM . La asignación se llevó a cabo utilizando el programa CARA

a los pares NH del esqueleto peptídico asignadas, y además, las señales correspondientes a grupos NH de cadenas laterales de asparagina y glutamina (rosa), y dos señales que corresponden al grupo indol de los dos triptofanos (verde).

Entre los 85 aminoácidos de la proteína las seis prolinas y el residuo N-terminal no dan señal en el $^1\text{H}^{15}\text{N}$ -HMQC, por lo que se esperan 78 señales correspondientes a los grupos amida del esqueleto. Se pudieron asignar 68 señales, quedando 6 sin asignar. Entre ellas 2 no corresponden al esqueleto peptídico de la proteína ya que no se observa el pico equivalente en el plano HN-N del HNCO (Figura 27), tratándose posiblemente de cadenas laterales de Lys o Arg. Las cuatro señales restantes corresponderían a HN amídicos del esqueleto peptídico pero no se pudieron asignar debido a que no se encontraron las correlaciones correspondientes en carbono. Quedan entonces seis residuos que no dan señal en el espectro. La ausencia de señales para estos grupos HN del esqueleto peptídico podría deberse a que los residuos están alojados en una región de movilidad intermedia de la proteína. Cuatro de estos residuos se localizan en el extremo N-terminal de DCL1-B, que concentra la casi totalidad de los residuos sin asignar. La ausencia de resonancias para la región N-terminal es común a los dominios dsRBD1 y 2 de HYL1, y probablemente refleje un motivo de movilidad intermedia común a todos estos dominios.

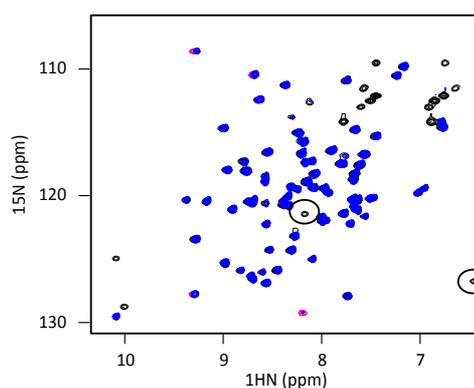


Figura 27: ^{15}N -HMQC y plano HN-N del HNCO de DCL1-B. Resaltadas en círculos se destacan las posibles señales de Arg y Lys

Una vez obtenidas las resonancias de los átomos del esqueleto peptídico, se realizó el cálculo del *Chemical Shift Index* (CSI) para establecer la estructura secundaria de la proteína. Como se puede observar en la figura 28, las estructuras secundarias obtenidas a partir de los desplazamientos químicos secundarios son las esperadas para un dominio de unión a ARN doble hebra, con la topología $\alpha\beta\beta\beta\alpha$.

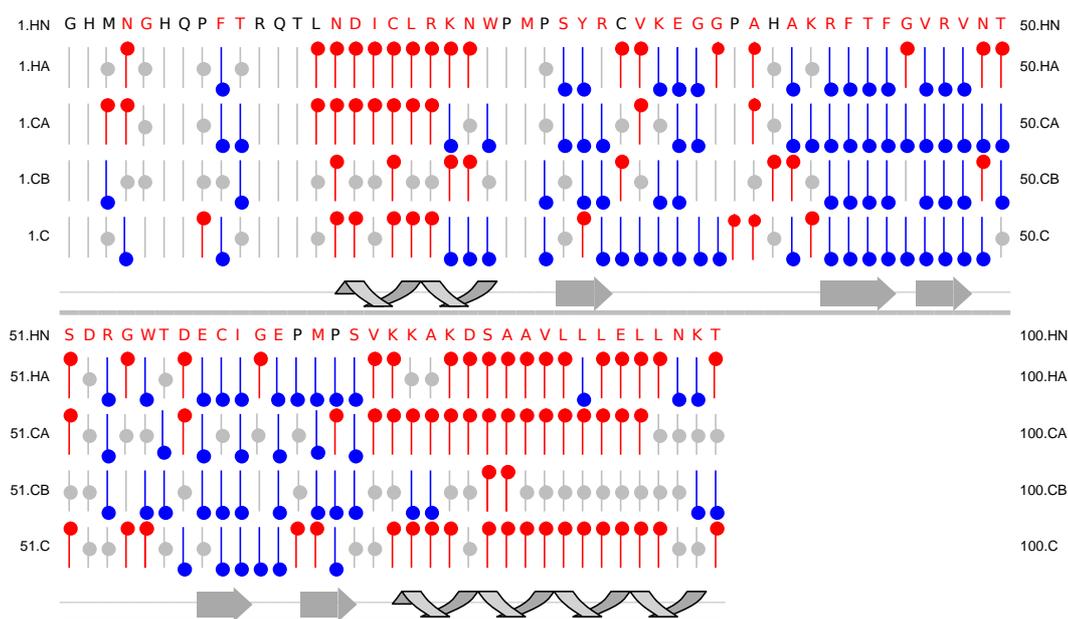


Figura 28: "Chemical shift index" (CSI) obtenido a partir del programa NMR View, utilizando la base de datos Wishart y los desplazamientos químicos de HA, CA, CB, y CO. En rojo se resaltan los residuos asignados

Luego de asignar las resonancias del esqueleto peptídico de la proteína. se realizó la asignación de las cadenas laterales a partir de espectros H(CC)(CO)NH y (H)CCH-TOCSY. La asignación se completó y confirmó utilizando los espectros 3D NOESY editados en ^{15}N y ^{13}C . La asignación completa de las cadenas laterales se depositó en BioMagResBank (entrada 18393) donde se incluyen 309 desplazamientos químicos correspondientes a ^{13}C , 67 a ^{15}N y 475 a ^1H , lo que

representa una asignación de $\sim 85\%$ del total. Las listas de los desplazamientos químicos obtenidos se utilizaron para asignar automáticamente los NOEs mediante los programas ATNOS-CANDID y CYANA.

4.1.5 CÁLCULO DE ESTRUCTURA EN SOLUCIÓN

A continuación se llevó a cabo el cálculo de la estructura en solución por RMN de DCL1-B. Para esto, se utilizaron restricciones de distancia provenientes de los NOEs, ángulos de torsión derivados del programa TALOS+ y los acoplamientos dipolares residuales de HN-N y HN-CO (ver sección 3.5.2.2). La estructura de la proteína incluye los residuos Asn 15 - Thr 85 de la construcción debido a que fuera de este rango, en el extremo N-terminal, se han podido identificar muy pocas resonancias. Las estadísticas de la estructura calculada se muestran en la tabla 1. La estructura fue depositada en el PDB (entrada 2LRS.pdb) (Figura 29).

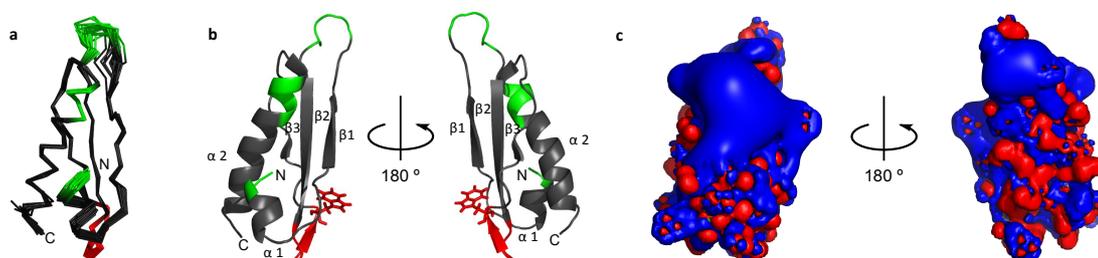


Figura 29: a) Representación del conjunto de las 20 estructuras de menor energía obtenidas del cálculo de estructura en solución. b) Estructura en solución de DCL1-B. Las regiones putativas de interacción con el ARN según el alineamiento de secuencias está coloreado en verde. La inserción entre $\beta 2$ y $\beta 3$ está representada en rojo. La cadena lateral del trp 55 se muestra como palos. c) Potencial electrostático del dominio. Se muestran isosuperficies a 3,5 V (azul) y -3,5 V (rojo). La orientación del dominio es la misma que en el panel b. Se observa una gran región positiva en las regiones putativas de unión a ARNdh.

<i>Nº de restricciones de distancia</i>	960
<i>Restricciones ambiguas</i>	428
<i>Intraresiduo</i>	212
<i>Secuenciales</i>	110
<i>Corto alcance</i>	18
<i>Largo alcance</i>	192
<i>Restricciones ambíguas</i>	179
<i>Restricciones de ángulos dihedros</i>	108 (52 ϕ , 52 ψ)
<i>RDCs</i>	
<i>Nº de restricciones de RDC</i>	112 (56HN-N, 56HN-CO)
<i>Tensor Aa</i>	$8,34 \times 10^{-4}$
<i>Tensor Ar</i>	$2,64 \times 10^{-4}$
χ^2	560,7
<i>Factor Q</i>	0,22
<i>Violaciones</i>	
<i>Distancia</i>	
<i>Raíz cuadrada media (rmsd)</i>	0,032 Å
<i>Mayores</i>	0,45 Å
<i>Dihedros</i>	
<i>rmsd</i>	0,87 °
<i>Mayores</i>	9,57°
<i>Desviación media de la geometría covalente ideal</i>	
<i>Longitud de enlace</i>	0,87 °
<i>Ángulo de enlace</i>	0,572 °
<i>Rmsd respecto a la estructura media (res 16-32, 42-83)</i>	
<i>Esqueleto peptídico</i>	0,29±0,07Å
<i>Átomos pesados</i>	0,80±0,10Å
<i>Análisis CING ROG (res 16-32, 42-83)</i>	
<i>Rojo</i>	12(20%)
<i>Naranja</i>	16(27%)
<i>Verde</i>	31 (53%)
<i>Análisis de Ramachandran (res 16-32, 42-83)</i>	
<i>Núcleo</i>	91,40%
<i>Permitidos</i>	8,60%
<i>Comunes y no permitidos</i>	0,00%

Tabla 1: Estadística para las 20 estructuras finales calculadas para DCL1-B con el programa ARIA/CNS

Es sabido que la unión al ARNdh por estos dominios es mediada por 3 regiones: la hélice $\alpha 1$ (R1), el bucle entre las láminas $\beta 1$ y $\beta 2$ (R2) y la región N-terminal de la hélice $\alpha 2$ (R3) ²¹². La estructura secundaria que conforman los

residuos que pertenecen a las regiones de interacción R2 y R3 parecen bien conservados en la estructura. Sin embargo, el primer motivo de interacción con el ARN (la hélice $\alpha 1$) es más corto que el de los dsRBDs canónicos. El hecho de que las resonancias del esqueleto peptídico del extremo N-terminal a partir de la Asn 15 no aparezcan en el espectro, sugiere que esta región no adoptaría una conformación estable. Esto se confirma con los bajos valores de ^1H - ^{15}N NOEs para las escasas señales que corresponden a los residuos Asn 4 – Asn 15 que están presentes en el espectro (Figura 30). La presencia de una hélice $\alpha 1$ mas corta en DCL1-B podría tener una implicación directa en la unión a ARNdh, ya que esta región juega un papel esencial en el reconocimiento de la estructura del sustrato ^{109,111,113,139,213} (Figura 31). La flexibilidad de esta región, podría aportarle a DCL1-B la posibilidad de reconocer moléculas de pre-miARNs con estructuras secundarias y terciarias sumamente heterogéneas, así como también ADNdh.

Otra característica interesante en la estructura de DCL1-B es la presencia de una inserción de seis residuos entre las láminas $\beta 2$ y $\beta 3$. Esta inserción está localizada en la cara opuesta a la de interacción con el ARN y posiblemente, no interfiera de manera directa en la especificidad y/o afinidad por los ácidos nucleicos. Dentro

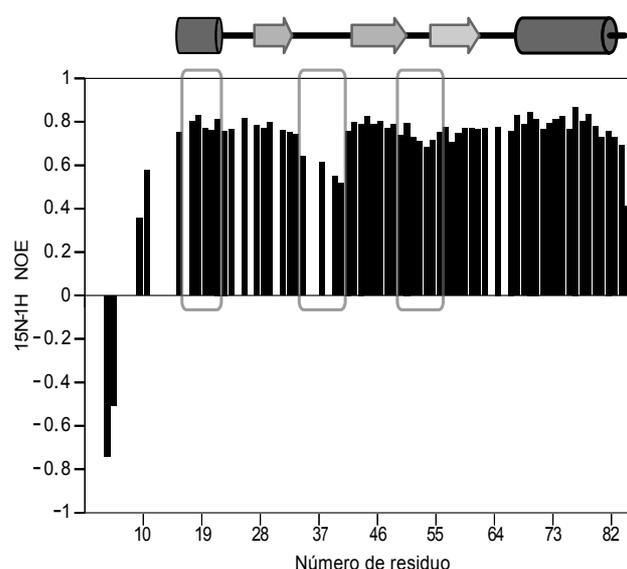


Figura 30: NOE Heteronuclear (^1H - ^{15}N) para DCL1-B. Los elementos de estructura secundaria están indicados en la parte superior. Las regiones recuadradas corresponden con la hélice $\alpha 1$, el bucle entre la lámina $\beta 1$ y $\beta 2$ y la inserción entre las láminas $\beta 2$ y $\beta 3$.

de la inserción, los grupos metilos de las Thr 43 y Thr 50 son parte de un bolsillo hidrofóbico que aloja a la cadena lateral del Trp 23, localizado en el bucle entre la hélice $\alpha 1$ y la lámina $\beta 2$. Se ha demostrado que la presencia de un residuo aromático en esta posición confiere estabilidad al plegamiento de los dsRBDs ¹⁷¹. Por este motivo, la extensión de este bucle con respecto a otros dsRBDs podría estabilizar a DCL1-B.

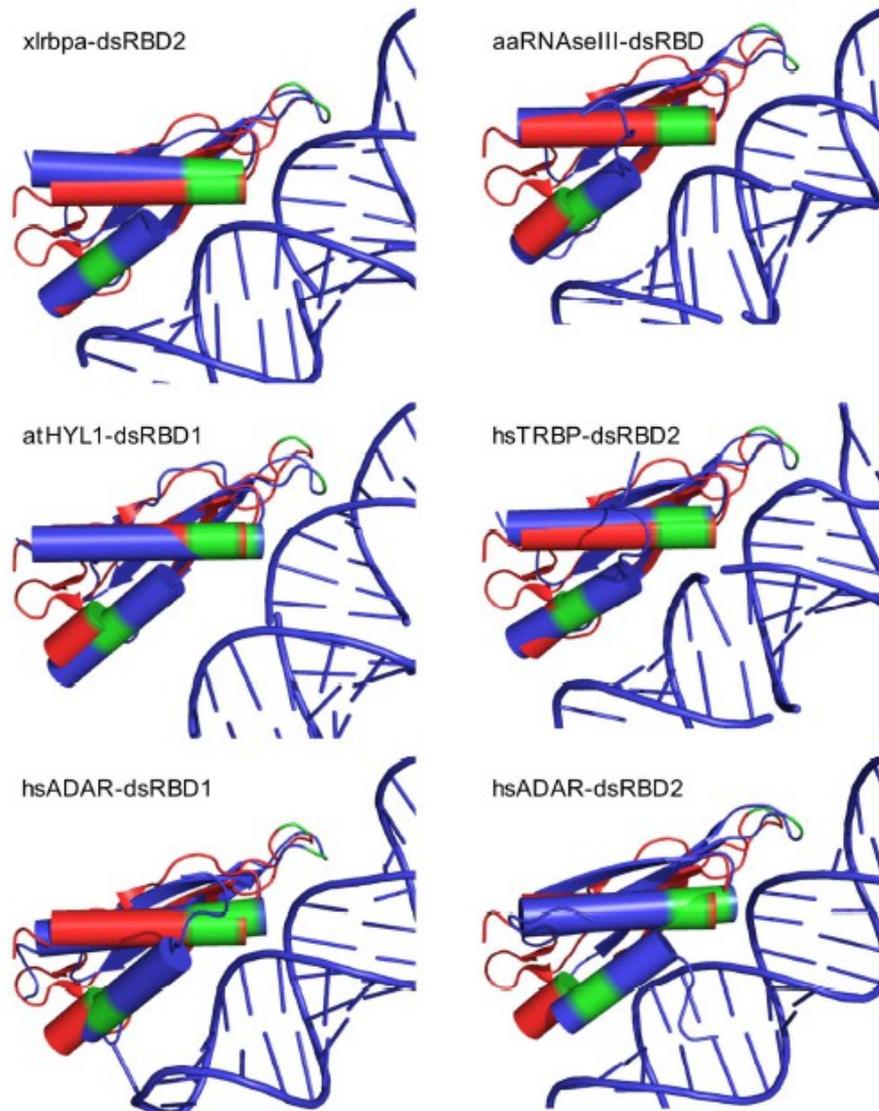


Figura 31: Superposición de estructuras de DCL1-B (rojo) con estructuras de complejos de distintos dsRBDs con ARNdh (azul): *Xlrpba-dsRBD2* (pdb 1di2); *aaRNaseIII-dsRBD* (pdb 2nug); *atHYL1-dsRBD1* (pdb 3adi); *hsTRBP-dsRBD2* (pdb 3adl); *hsADAR-dsRBD1* (pdb 2l3c); *hsADAR-dsRBD2* (pdb 2l2k). Las regiones de unión a ARN están coloreadas en verde.

Otra particularidad de este dominio es el hecho de que la cadena lateral del Trp 55, localizado dentro de la inserción, sobresalga en la cara opuesta a la superficie putativa de unión al sustrato. Este residuo de Trp está altamente conservado entre los DCL1-B de diferentes plantas, pero no está presente en las secuencias de DCL3-B o DCL4-B (Figura 32). La presencia de un residuo aromático voluminoso en la superficie de la proteína ²¹⁴ sugiere una superficie de interacción proteína-proteína que podría resultar importante para la función de DCL1 en el contexto del complejo de procesamiento de miARNs.

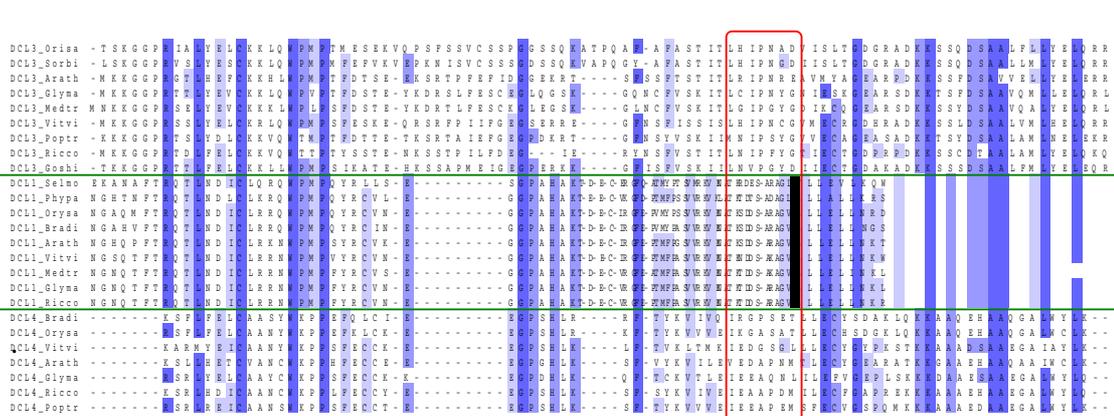


Figura 32: Alineamiento de secuencias de DCL-B de plantas. Las secuencias de los distintos tipos de DCLs están separadas con líneas verdes. Los residuos están coloreados en distintos tonos de azul de acuerdo al grado de conservación. La región recuadrada corresponde a la inserción entre las láminas 2 y 3. Los Trp 55 (1877 dentro de la secuencia completa de DCL1) está resaltada con fondo negro.

Por otro lado, durante la asignación de las resonancias del ¹H¹⁵N-HMQC, se pudo observar la presencia de dos señales por residuo en algunos casos (Figura 33). Estas dos señales fueron asignadas al mismo residuo debido a que las resonancias del CA y CA-1 son idénticas. Este es el caso de los residuos Arg 47-Val 48 por un lado y los residuos Ala 74, Leu 77, Leu 79, Leu 81 y Leu 82, que estarían formando parte de la segunda lámina β y del extremo C-terminal de la hélice α2, respectivamente, según el cálculo del “Chemical Shift Index” y del mapeo en la

estructura tridimensional en solución (Figura 28 y 33). La presencia de señales dobles, sugiere la existencia de conformaciones alternativas en una región de la proteína. La población de estas conformaciones sería variable con el tiempo, lo que pudo evidenciarse tras la desaparición de una de las dos señales del HMQC adquirido al comienzo y al finalizar los experimentos de triple resonancia para la asignación (Figura 33).

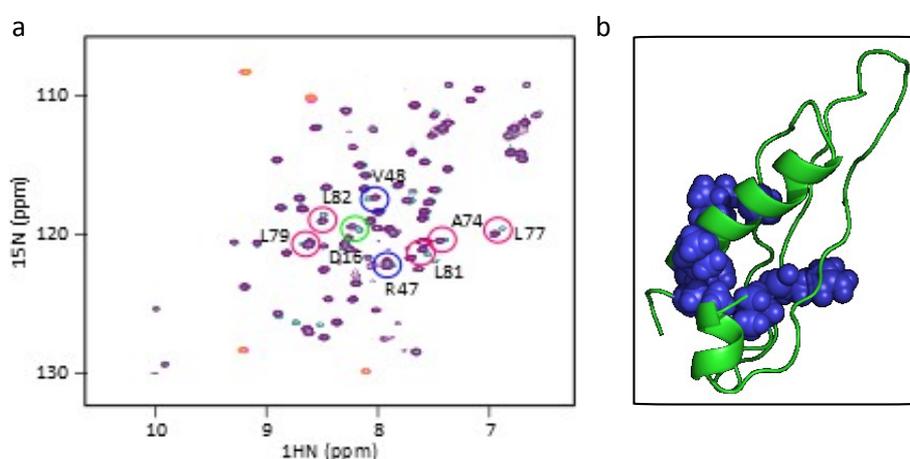


Figura 33: a) Superposición de ^{15}N -HMQCs antes (celeste) y después (violeta) de la adquisición de los experimentos de triple resonancia. Resaltados en círculos se destacan los residuos que dan señales dobles al comienzo de la adquisición. En azul se resaltan los aminoácidos que corresponden a la lámina 2; en rosa los que corresponden al extremo C_terminal de la hélice 2 y en verde el Asp 16. b) Mapeo de las señales dobles que aparecen inicialmente en los espectros $^1\text{H}^{15}\text{N}$ -HMQC sobre la estructura de DCL1-B.

4.1.5.1 CÁLCULO DE ESTRUCTURA POR RDC

Tal como se mencionó previamente, para la determinación de la estructura en solución, se utilizaron como restricción adicional las medidas de acoplamiento dipolares residuales. Las medidas de RDC, permiten calcular la orientación relativa de pares de átomos respecto del marco proteico, por lo que utilizando solamente los acoplamiento N-HN y CO-HN, es posible calcular el

plegamiento de una proteína de forma rápida y precisa, sin necesidad de pasar por las etapas que más tiempo consumen: la asignación de cadenas laterales y de los espectros NOESY ²¹⁵. En la Figura 34 se pueden observar las 10 estructuras de menor energía de DCL1-B obtenidas con el programa Rosetta a partir de los RDC.

Una forma de determinar la calidad de la estructura obtenida por RDC es analizar la correlación entre los valores de RDC medidos respecto a los RDC calculados a partir de la estructura. Dicho ajuste, arroja un parámetro de calidad denominado factor Q. Este factor puede tener valores de 0 a 1, siendo los valores más bajos los que manifiestan mejor

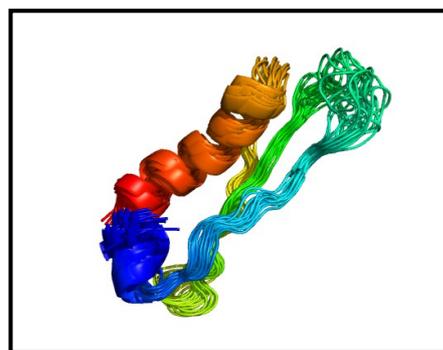


Figura 34: Representación de las diez estructuras de menor energía obtenidas usando Rosetta con parámetros de RDCs N-HN y CO-NH.

correlación y por lo tanto, mayor calidad del cálculo de la estructura. Los gráficos de correlación entre los valores medidos y calculados para la estructura de menor energía de DCL1-B muestran en general una buena correspondencia, excepto en algunas regiones particulares. El valor del factor Q para los RDC N-HN es de 0,193, mientras que para el par CO-NH es de 0,281. Respecto a los residuos que presentan baja correlación, se observó que existe una sola región que muestra baja correlación en el ajuste para ambos pares (N-NH y CO-NH). Esta región incluye a los aminoácidos comprendidos entre la Asn 49 y la Thr 56 y corresponde al bucle β 2- β 3 que contiene la inserción de seis aminoácidos que es característica de DCL1 (Figura 35). Probablemente esta baja correlación se deba a que es una región dinámica de la proteína. Si bien esto no se refleja en el experimento de NOE heteronuclear, sí se puede evidenciar a partir de medidas de relajación (ver más adelante, sección 4.1.8.2). Existen otras regiones donde se

encuentra una baja correlación en los acoplamientos CO-NH. Éstas corresponden al primer bucle, donde la falta de correlación puede deberse a que es una región de alta movilidad y a la región N-terminal de la segunda hélice.

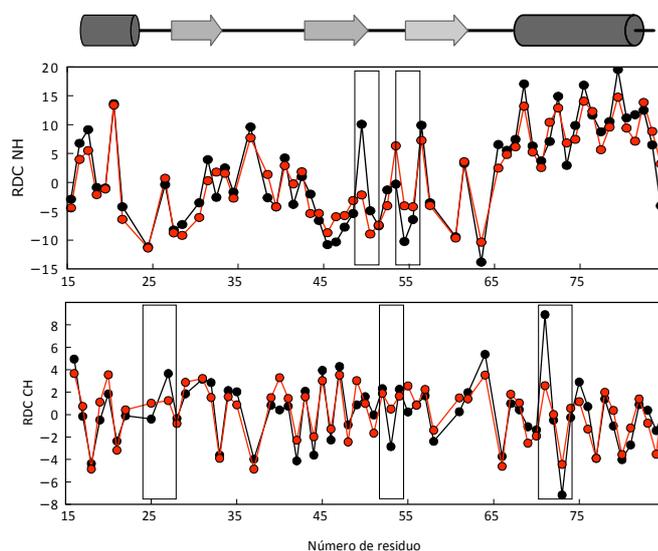


Figura 35: Correlación entre los RDC medidos y los RDC calculados a partir de la estructura de la proteína calculada con Rosetta. Arriba se muestra el RDC observado (negro) y calculado (rojo) para el par N-NH correspondientes a la estructura de menor energía. Abajo, se presenta el mismo gráfico para el par CO-NH. La numeración corresponde al fragmento calculado (residuos 15-85).

Al comparar la estructura obtenida por medidas de RDC respecto a la estructura obtenida por restricciones de distancia (NOEs), se pudo notar que si bien ambas estructuras son similares, existe una diferencia importante en la orientación que adopta la primer hélice (Figura 36). Esta diferencia puede deberse a la falta de restricciones aportadas por las cadenas laterales en el cálculo realizado en CS-Rosetta. Por otro lado, luego de realizar un análisis de correlación de los RDC medidos con respecto a los RDC calculados a partir de la estructura obtenida de los NOEs, se observó que si bien mejora la correlación

respecto a la de la estructura calculada por Rosetta, hay algunas regiones donde existe un defasaje. En este caso, los ajustes más bajos se observan para la región que corresponde al bucle ubicado entre la lámina $\beta 1$ y $\beta 2$. Esta falta de correlación se da tanto para los pares N-NH como para CO-NH. Esto podría deberse a que en esta región pesa más el valor de NOE que la restricción del RDC para el cálculo. La presencia de movilidad puede afectar la correlación entre los RDC observados y calculados. Por este motivo, la baja

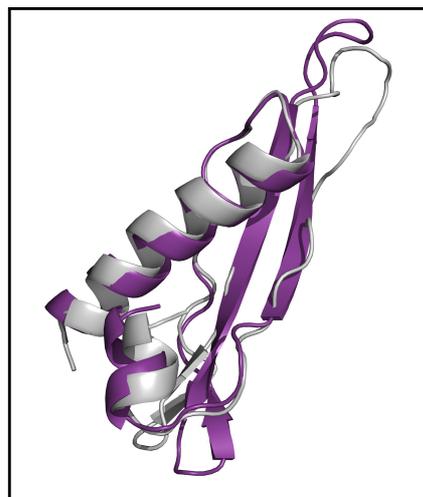


Figura 36: Comparación de las estructuras obtenidas a partir de restricciones de RDC (gris) y de distancia a partir de los NOE (violeta).

correlación, podría deberse a la presencia de cierta flexibilidad en esa región, lo que se ve manifestado por los bajos valores de NOE heteronuclear (Figura 30). Al igual de lo que ocurre con el ajuste de la estructura obtenida por RDC, la región N-terminal de la hélice $\alpha 2$ presenta valores de RDC que no ajustan correctamente (Figura 37). Sin embargo, a diferencia de lo que se observa para el ajuste con la estructura obtenida por Rosetta, no existen problemas de correlación en el bucle $\beta 2$ - $\beta 3$, lo que indicaría que las restricciones de cadenas laterales aportadas por los NOEs en el cálculo de estructura aumentarían la precisión en el cálculo de la estructura en esta región.

A pesar de las diferencias observadas entre las estructuras obtenidas por medidas de RDC y por NOEs, es importante destacar que el cálculo a partir de CS-Rosetta brinda de una manera rápida una estructura que puede resultar sumamente útil para realizar estudios que requieran conocer únicamente el

plegamiento de la proteína y no detalles sobre la orientación de cadenas laterales.

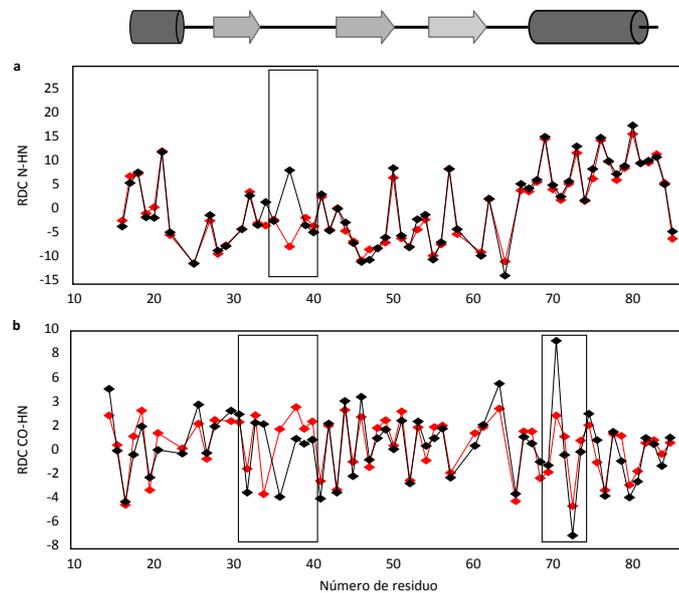


Figura 37: Correlación entre los valores de RDC medidos y los calculados para la estructura 2LRS.pdb. Arriba se muestra el RDC observado (negro) y calculado (rojo) para el par N-NH correspondientes a la estructura. Abajo, se presenta el mismo gráfico para el par CO-NH. La numeración corresponde al fragmento calculado (residuos 15-85).

4.1.6 INTERACCIÓN PROTEÍNA-PROTEÍNA

Como se mencionó en la sección 1.5, los dsRBDs tienen otras funciones además de la de reconocer ácidos nucleicos. Una de ellas involucra la capacidad de interactuar con otras proteínas. En este sentido, se ha reportado que los dsRBD del extremo C-terminal de DCL1 interactúan con los dsRBDs de HYL1²¹⁶. Con el objetivo de evaluar esta posibilidad, se decidió investigar si DCL1-B participa en esta interacción *in vitro*. Para esto, se realizó una titulación de DCL1-B marcada con ¹⁵N con los dos dsRBDs de la proteína HYL1 (HYL1-AB) sin marca

isotópica. Al analizar los espectros, se observó que el $^1\text{H}^{15}\text{N}$ -HSQC correspondiente a DCL1-B no muestra ningún cambio significativo luego de la adición de los dsRBDs de HYL1, lo que estaría indicando que no existe interacción entre ambas proteínas (Figura 38).

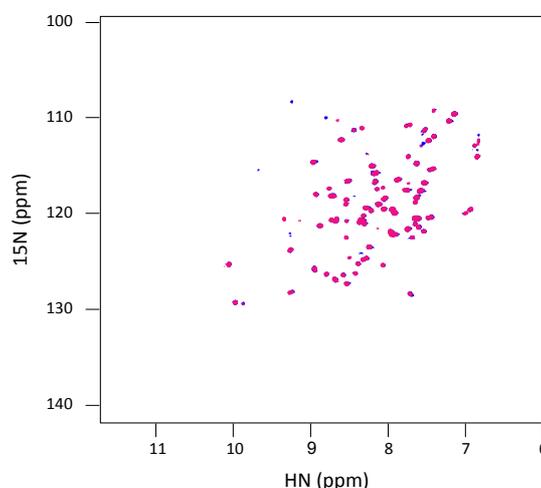


Figura 38: Espectro ^1H - ^{15}N HSQC de ^{15}N -DCL1-B libre (azul) y en presencia de un equivalente de HYL1-AB sin marca (rosa). No se observan cambios significativos en el espectro de DCL1-B luego de la adición de HYL1-AB.

Para evaluar si la falta de interacción se debía a la ausencia del primer dsRBD de DCL1, decidimos realizar el mismo experimento utilizando una construcción de ambos dsRBDs de DCL1 (DCL1-AB) con marca isotópica ^{15}N junto con HYL1-AB sin marca y por otro lado, de HYL1-AB ^{15}N con DCL1-AB sin marca. En ninguno de los dos casos se observaron cambios en el espectro luego de la adición de la proteína sin marca, lo que estaría confirmando que ni DCL1-B ni DCL1-AB interaccionarían con HYL1-AB, al menos, de forma directa (Figura 39).

4.1.7 UNIÓN DE DCL1-B A ÁCIDOS NUCLEICOS

Con el objetivo de investigar el reconocimiento de ARN por DCL1-B, se realizaron medidas de anisotropía de fluorescencia. Las determinaciones de las afinidades se hicieron midiendo el cambio en la anisotropía de fluorescencia que se produce al formarse el complejo ARN-proteína.

El estudio de la interacción de DCL1-B con ARNdh se realizó utilizando dos

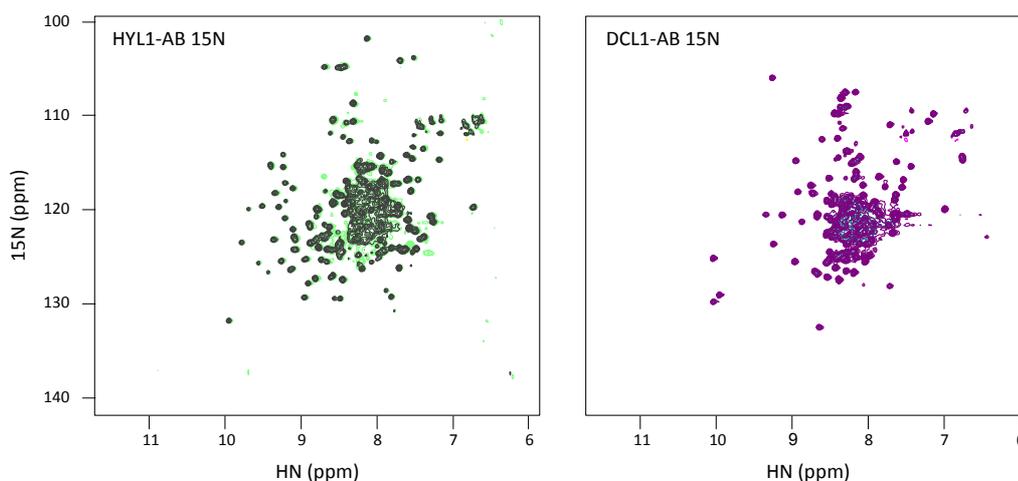


Figura 39: Espectros $^1\text{H}^{15}\text{N}$ HSQC de ^{15}N -HYL1-AB libre (verde) superpuesto con el de ^{15}N -HYL1-AB en presencia de un equivalente de DCL1-AB sin marca (negro) y de ^{15}N -DCL1-AB libre (celeste) superpuesto con el de ^{15}N -DCL1-AB en presencia de un equivalente de HYL1-AB sin marca (violeta). No se observan cambios significativos ninguno de los espectros luego de la adición de proteína.

fragmentos distintos del precursor miR172 (pri-miR172); uno que corresponde al miARN/miARN* y otro correspondiente a la región del tallo inferior del miR172 (pre-miARNs). Recientemente se ha demostrado que esta última región es esencial para el registro del primer corte en el procesamiento del pri-miARN en plantas y es menos estructurada que la región del miARN/miARN* ¹⁰¹.

La determinación de las constantes de afinidad se llevó a cabo según se describe en la sección 3.8. Tras realizar los ajustes, observamos que la afinidad de DCL1-B por ambas construcciones de ARNdH es similar, con K_D , **pre-miARNs** = 350 ± 20 nM y K_D , **pri-miR172** = 810 ± 180 nM. Estos resultados indican que es improbable que este dominio sea capaz de reconocer características particulares dentro de la estructura del precursor. Sin embargo, el hecho de que la afinidad hacia el pre-miR172s sea algo mayor que hacia el *duplex* miARN/miARN*, sugiere que DCL1-B tendría alguna preferencia por el ARNdH que presenta imperfecciones (Figura 40). El incremento de la fuerza iónica en la solución

amortiguadora disminuye la fuerza de la interacción, lo que evidencia que la unión al ARN es mediada en gran parte por un componente electrostático (Figura 40). El cálculo del potencial electrostático de DCL1-B apoya esta observación, debido que la región putativa de interacción con el ARN, muestra un gran potencial positivo (Figura 29).

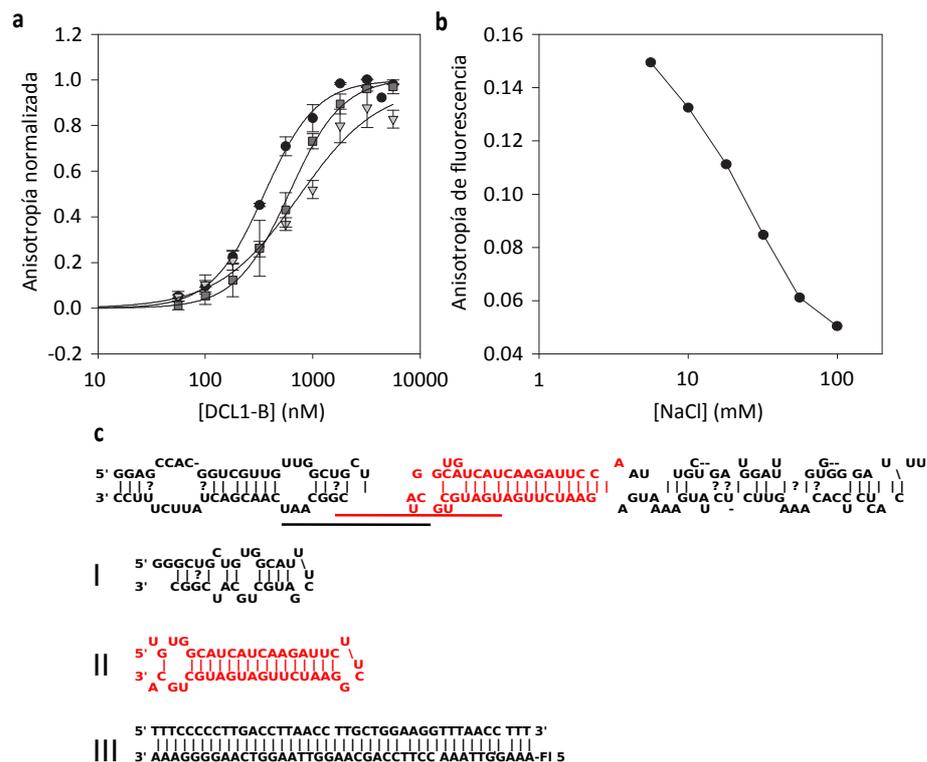


Figura 40: Interacción de DCL1-B con ARNdh y ADN seguida por anisotropía de fluorescencia. a) Titulación de ácidos nucleicos marcados fluorescentemente con DCL1-B. Cada punto corresponde a un promedio de tres (miR172-Is, círculos) o dos (miR172, triángulos; ADN cuadrados) experimentos independientes. Los datos para el miR172-Is y ADNdh fueron ajustados a la ecuación de Hill (coeficiente de Hill igual a 1,8 en ambos casos). b) Dependencia de la fuerza iónica en la unión de DCL1-B con ARNdh en una solución de 50 nM del miR172-Is marcado fluorescentemente y 3,2 μ M de DCL1-B. La reducción de la anisotropía de fluorescencia indica la disociación del complejo a elevada fuerza iónica. Las abscisas están graficadas en escala logarítmica. c) Construcciones de ácidos nucleicos utilizadas. Arriba, pri-miR172, el miARN/miARN* está coloreado en rojo. El constructo I corresponde a la región del tallo inferior, el II a la región del miARN y el III a la secuencia de ADNdh.

A modo de control, se decidió evaluar la unión de DCL1-B a ADNdh utilizando el mismo tipo de ensayo. De manera inesperada, observamos que DCL1-B une ADNdh con una afinidad similar a la que se une al ARNdh, con valor de $K_D = 600 \pm 50 \text{ nM}$ (Figura 40). En un experimento control realizado con el primer dsRBD de HYL1 (HYL1-A), un dsRBD canónico, encontramos una gran diferencia de afinidad entre la unión a ARNdh y ADNdh (Figura 41). Para este caso, el valor de K_D obtenido para el pre-miR172-Is es de 310 nM, mientras que el de ADN no pudo ser estimado debido a que la K_D presenta un valor mayor a 20 μM . El hecho de que DCL1-B se una al ADNdh con alta afinidad es una característica poco común entre los dominios de unión a ARNdh, debido a que estos dominios se caracterizan por su capacidad de discriminar entre diferentes tipos de ácidos nucleicos (ARN y ADN simple hebra y ADNdh de ARNdh). Una posible causa de la especificidad dual de este dominio, podría ser a la naturaleza dinámica de los residuos localizados hacia el extremo N-terminal de la hélice α_1 de DCL1-B, acompañada de la componente electrostática de la interacción proteína-ácido nucleico.

Con el propósito de conocer cuales son los determinantes que este dominio utiliza para el reconocimiento de los ácidos nucleicos, se llevó a cabo el mapeo de las regiones de interacción a partir de titulaciones de la construcción del pre-miR172Is y ADNdh sobre muestras de proteína marcadas con ^{15}N siguiendo los

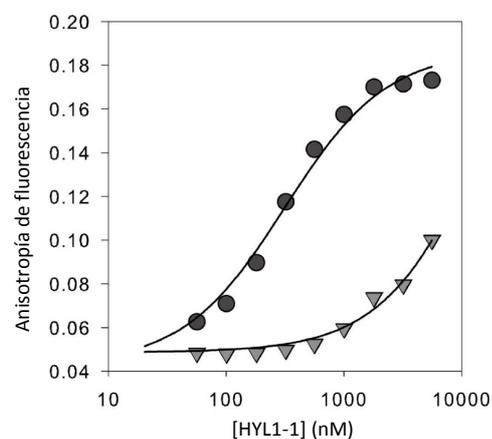


Figura 41: Titulación por anisotropía de fluorescencia de miR172-Is (círculos) y ADNdh (triángulos) con HYL1. Las líneas sólidas muestran los ajustes de los valores experimentales. La abscisa está graficada en escala logarítmica.

cambios en espectros $^1\text{H}^{15}\text{N}$ -HSQC. En primer lugar, se llevó a cabo la titulación de proteína sobre una muestra de ARN. Los espectros obtenidos a relaciones ARN:proteína sub-estequiométricas muestran señales anchas pero no se observa un conjunto de señales nuevas respecto al espectro de la proteína libre. Luego de la adición de exceso de proteína (relación ARN:proteína 1:1,5), las señales correspondientes a la proteína libre no reaparecen, indicando que la estequiometría de unión proteína:ARN en estas condiciones es mayor a 1:1 (Figura 42). Al no poder identificar las regiones de DCL1-B involucradas en la interacción con el ARN mediante el experimento anterior, se realizó una titulación de ARN sobre la proteína. En este caso, se observó que la mayoría de las señales decaían en intensidad, aún a muy baja relación ARN:proteína (1:100), sin cambios significativos en los desplazamientos químicos. Al mismo tiempo, las señales ^1H los grupos imino del ARN eran claramente visibles, demostrando que la disminución en la intensidad de las señales de la proteína no se debía a la formación de complejos de gran tamaño o agregados ya que esto produciría un ensanchamiento de las señales del ARN. Este comportamiento, sugiere que la proteína en el complejo con ARN se encuentra en un régimen de intercambio intermedio, y que se estaría uniendo al ARN adoptando múltiples conformaciones. A pesar de que el espectro muestra una disminución generalizada de la intensidad de las señales, aún a baja relación ARN:proteína, el análisis de los perfiles de la intensidad de las señales muestran que el bucle $\beta 2$ - $\beta 3$ y la región C-terminal de la hélice $\alpha 2$ son menos afectadas que el resto de la proteína (Figura 42 y Figura 43), lo que indica que no participan en la unión. Este experimento demuestra que DCL1-B une ARN de una forma similar a la que lo hacen otros dsRBD canónicos.

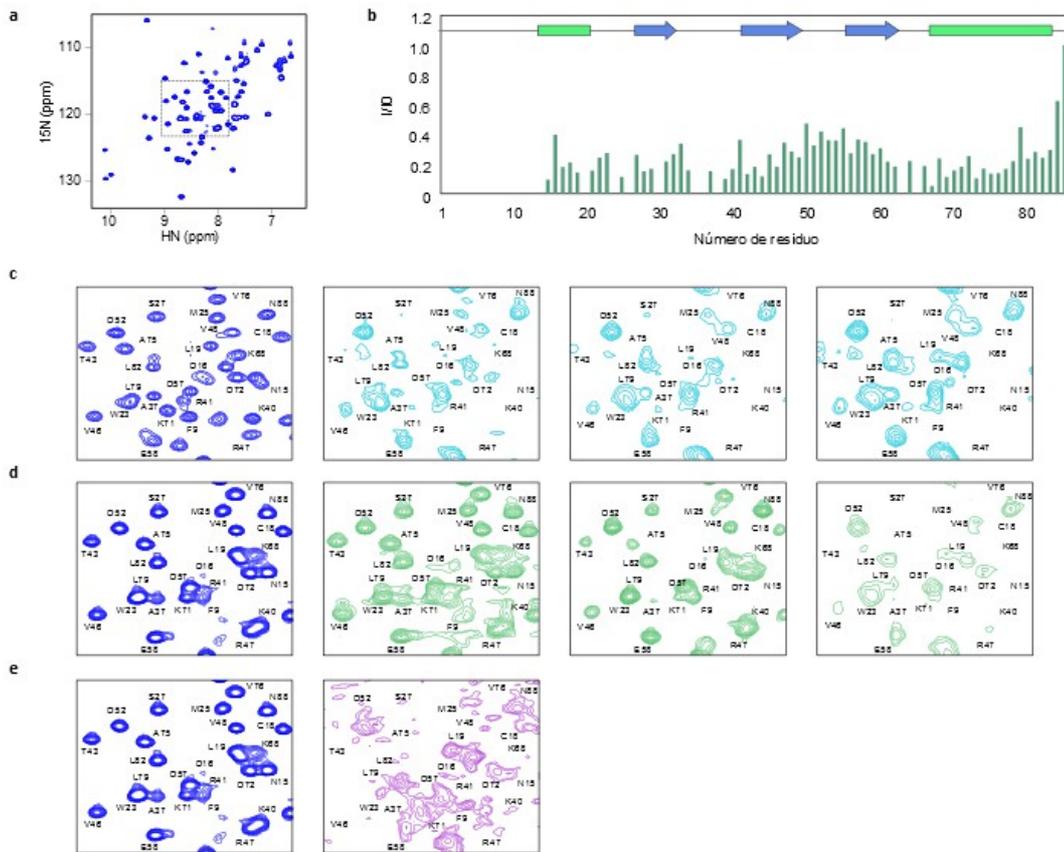


Figura 42: Caracterización de la interacción de DCL1-B con ácidos nucleicos por RMN. a) Espectro de DCL1-B libre. El rectángulo discontinuo muestra la región que está ampliada en los paneles c, d y e. b) Relación de la intensidad de la señal vs secuencia para los espectros de DCL1-B en presencia de 0,05 equivalentes de pre-miR172ls. c) Titulación de ARN con DCL1-B seguida por RMN. Los espectros corresponden a (de izquierda a derecha) proteína libre, y a relaciones de ARN:proteína de 2, 1 y 0,66, respectivamente. d) Titulación de DCL1-B con ARN seguida por RMN. Los espectros corresponden a (de izquierda a derecha) proteína libre, 0,01; 0,05; y 0,1 equivalentes de pre-miR172ls, respectivamente. e) Espectro de RMN de DCL1-B libre y DCL1-B en presencia de 0,05 equivalentes de ADNdH.

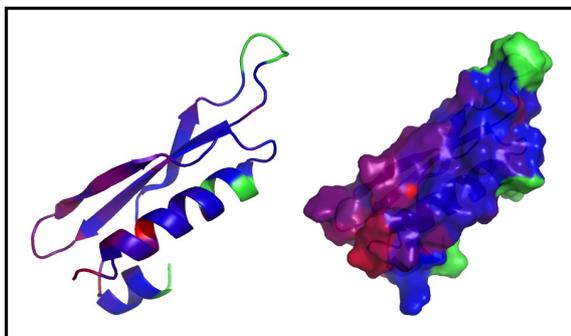


Figura 43: Mapeo de la interacción de DCL1-B con el pre-miR172ls sobre la estructura utilizando los valores de I/O que se muestran en la figura Figura 42. El gradiente de colores de rojo a azul indican las regiones menos afectadas a las más afectadas. Las regiones de interacción canónicas están coloreadas en verde.

4.1.8 MEDIDAS DE DINÁMICA POR RESONANCIA MAGNÉTICA NUCLEAR

4.1.8.1 ANÁLISIS DE LOS DATOS DE RELAJACIÓN

Se llevó a cabo un estudio de la dinámica del esqueleto peptídico de DCL1-B en la escala ps – ns midiendo las velocidades de relajación longitudinal y transversal (R1 y R2) de los átomos de ^{15}N y el NOE heteronuclear ^1H - ^{15}N en estado estacionario. Los fundamentos teóricos y los detalles técnicos de los experimentos realizados se describen en las secciones 3.5.1.5 y 3.5.2.5, respectivamente.

Los parámetros de relajación (R1, R2 y NOE), se obtienen a partir de una serie de espectros ^1H ^{15}N -HSQC relacionados, con diferentes tiempos de relajación o de evolución. La disminución de la intensidad de las señales al incrementarse los tiempos de retardo reflejan los decaimientos correspondientes a los procesos de relajación longitudinal (R1) y relajación transversal (R2). En la Figura 44 se muestran los perfiles de R1, R2 y NOE heteronuclear obtenidos para DCL1-B. Para el análisis, no se tuvieron en cuenta aquellos residuos que presentan señales total o parcialmente solapadas en los espectros ^1H ^{15}N -HSQC. Tampoco se consideraron los valores de R1 y R2 de aquellos residuos para los cuales se obtuvo un ajuste visualmente pobre de los datos experimentales. Los valores promedio de R1 y R2 son $1,87 \pm 0,16 \text{ s}^{-1}$ y $14,32 \pm 2,34 \text{ s}^{-1}$; respectivamente, mientras que el valor promedio de los NOEhn es de $0,69 \pm 0,04$. Estos valores promedio, se corresponden bien con los parámetros calculados para otros dsRBDs ²¹⁷. El análisis de los datos de NOEhn muestra que las regiones de mayor movilidad de DCL1-B serían los extremos N y C-terminal y el bucle β 1- β 2. La región correspondiente a el bucle β 2- β 3, donde está localizada la inserción de seis aminoácidos, presenta una disminución leve de los valores de NOE, lo que estaría indicando menor movilidad respecto a el otro bucle localizado entre dos

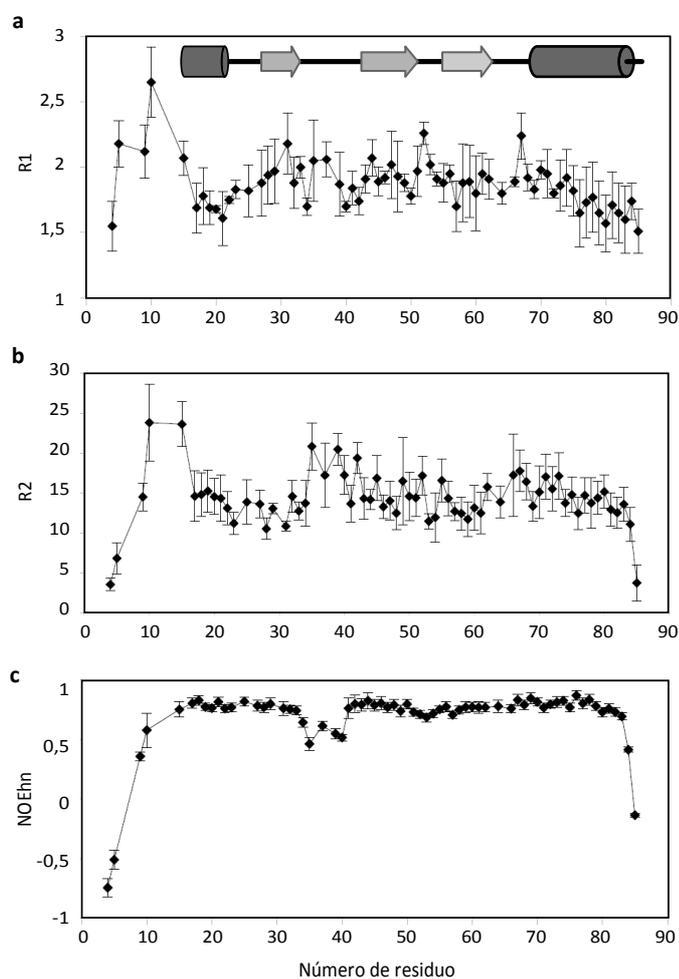
láminas β .

Figura 44: Perfiles de a) R_1 , b) R_2 , c) NOE heteronuclear y d) R_2/R_1 , en función de la secuencia proteica de DCL1-B.

4.1.8.2 ANÁLISIS "MODEL FREE"

El análisis de los parámetros de relajación mediante el formalismo *Model Free*, implementado en el programa Tensor v2.0¹⁷⁹, arrojó un valor de tiempo de correlación rotacional (τ_c) de $7,93 \pm 0,10$ ns. Este valor se encuentra dentro del rango de los τ_c calculados para otros dsRBDs²¹⁷⁻²¹⁹. El modelo de movilidad que mejor ajustó a los datos de R_1 y R_2 ingresados fue el modelo anisotrópico, con

los siguientes valores para el tensor de difusión:

$$D_x = 0,0179 \pm 0,001 \text{ ns}^{-1}$$

$$D_y = 0,0181 \pm 0,001 \text{ ns}^{-1}$$

$$D_z = 0,0261 \pm 0,001 \text{ ns}^{-1}$$

En la figura 45 se muestra el tensor de difusión rotacional correspondiente a DCL1-B. El cálculo muestra una buena correspondencia entre el tensor de difusión rotacional calculado a partir de los datos de relajación, el tensor de inercia y la geometría de la molécula.

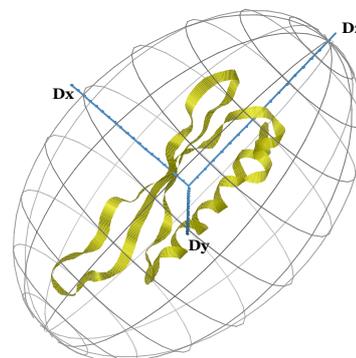


Figura 45: Tensor de difusión rotacional de DCL1-B calculado mediante el programa TENSOR2.

Los perfiles de S^2 graficados en función de la secuencia proteica se muestran en la figura Figura 46. A primera vista, se observa que los valores de S^2 son notablemente altos para la mayoría de los residuos de la proteína, alcanzando un valor promedio de $0,82 \pm 0,09$. Es importante destacar que cuando los valores de S^2 son menores a 0,6 se puede asumir que la dinámica en la escala de ps-ns es significativa. En el caso de DCL1-B, se destacan cuatro regiones que presentan mayor movilidad, con valores de S^2 bajos. Tres de ellas coinciden con las regiones donde el dominio interacciona con el ARNdh. La primera región corresponde a la hélice α_1 , la segunda al bucle 2 y la tercera al bucle 4 y parte de la hélice α_2 (Figura 46).

El hecho de que se observe mayor movilidad para los residuos que participan del reconocimiento del ARN coincide con lo reportado para otros dsRBDs. La cuarta región flexible pertenece a la región de la inserción de seis aminoácidos localizada entre las láminas β_2 - β_3 . Probablemente la flexibilidad de esta región sea necesaria para poder llevar a cabo interacciones con otras

proteínas mediadas por el Trp 55. Adicionalmente se observa que los NH amídicos que ajustaron a un modelo que incluye una contribución de intercambio (movimientos en la escala de ms a μ s) corresponden también a estos residuos localizados en regiones de alta movilidad (Figura 47).

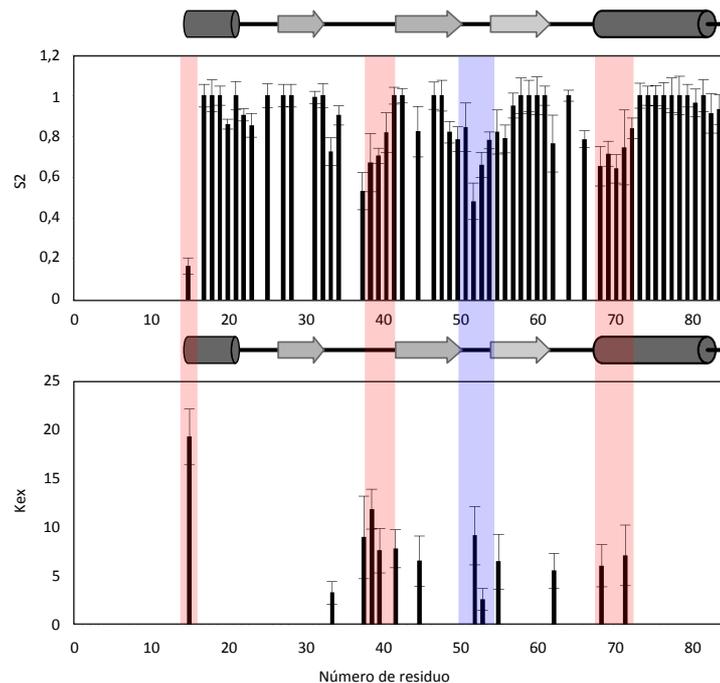


Figura 46: Parámetros dinámicos obtenidos en TENSOR2. En la figura superior se muestra el S^2 en función de la secuencia proteica y en el gráfico inferior la K de intercambio (K_{ex}) para los residuos que corresponden. En rojo se destacan las regiones de interacción con el ARNd y en azul la inserción en el bucle $\beta 2-\beta 3$.

4.1.9 LOCALIZACIÓN DE DCL1-B *IN VIVO*

Numerosos trabajos han reportado que la proteína DCL1 de plantas tiene localización nuclear, sin embargo, hasta el momento, no se ha evidenciado la presencia de una señal de localización nuclear (NLS) putativa en su secuencia. Existen estudios que han demostrado que existen algunas NLS que difieren de las

clásicas señales mono o bipartitas, indicando que las señales de direccionamiento al núcleo tienen naturaleza divergente. Un ejemplo de estas señales atípicas lo constituyen los dsRBDs de algunas proteínas como el tercer dsRBD de ADAR1^{128,129}, el dsRBD extendido de DICER de *S. pombe*¹³⁰ y el dsRBD de DICER humana¹³¹. Esta función de localización no es común a todos los dsRBDs sino que es una característica particular de algunos de ellos, la cual no puede ser predicha a partir de la secuencia¹³¹.

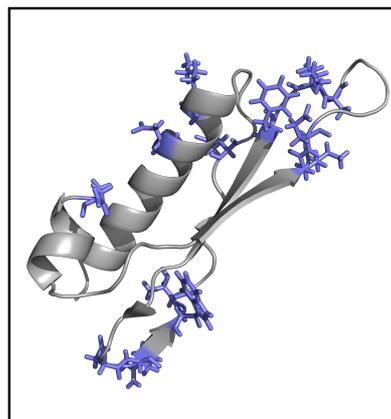


Figura 47: Mapeo de los residuos que presentan intercambio sobre la estructura de DCL1-B.

Teniendo en cuenta que DCL1-B representa a un dsRBD no canónico y que posiblemente su función no esté relacionada con el reconocimiento específico del sustrato, se decidió evaluar si este dominio era capaz de dirigirse al núcleo utilizando una proteína reportera. Con este propósito, se transformaron plantas de *Arabidopsis thaliana* cv. *Columbia* con construcciones de los dsRBDs de DCL1 fusionados a la proteína verde fluorescente de *Aequorea victoria* (GFP). Como control negativo, se utilizó el vector vacío pCHF3_CS052. A primera vista se puede evidenciar que tal como se esperaba por la presencia del promotor CaMV35S, la expresión de los dominios fusionados a GFP es constitutiva en todas las células de la raíz (Figura 48). Al analizar las imágenes obtenidas, se encontraron diferencias en la localización de la proteína reportera fusionada a ambos dominios. Mientras que DCL1-A se localiza de manera difusa en el interior de las células radiculares, DCL1-B presenta localización nuclear. Esto sugiere que el segundo dsRBD de DCL1 podría actuar como una potencial señal

de localización nuclear (NLS) para DCL1. La diferencia de localización entre los dos dsRBDs de DCL1, podría indicar que ambos dominios son funcionalmente divergentes, aunque será necesario realizar estudios más exhaustivos para confirmar esta función de DCL1-B.

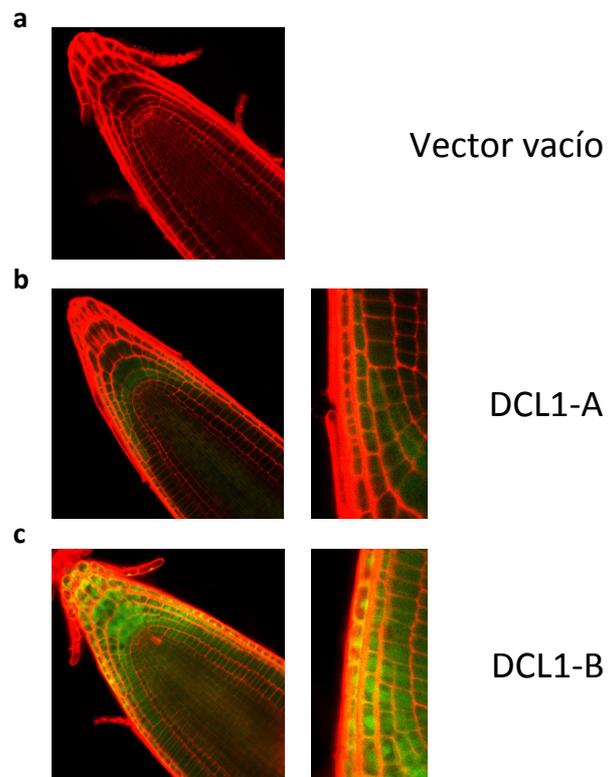


Figura 48: Microscopía confocal de raíces de plantas de *A. thaliana* transformadas con los dominios dsRBD de DCL1 fusionados a GFP (verde). a) Vector pCHF3_CS052. b) DCL1-A, c) DCL1-B. A la derecha de b y c se muestra una ampliación de la imagen. La pared de las células vegetales se tiñó con yoduro de propidio (rojo)

4.2 DISCUSIÓN 1

El dominio DCL1-B es esencial para el funcionamiento correcto de DCL1. Se ha demostrado que la interrupción de la proteína en este dominio (mutante *dcl1-9*) conduce a una disminución general de los niveles de miARN de las plantas, lo que se evidencia por anomalías fenotípicas²²⁰. Se ha sugerido que el segundo dsRBD de DCL1 es necesario para el reconocimiento correcto de la estructura de los precursores y para la localización correcta de los sitios activos ARNasa III sobre el precursor³⁷. En base a estos datos, trabajamos bajo la hipótesis de que DCL1-B podría unirse reconociendo algún motivo de estructura de los precursores. Sin embargo, este trabajo observamos que si bien DCL1-B es capaz de unirse al precursor de miARN, carece de especificidad hacia alguna región específica del precursor, lo que se demostró a partir de los espectros de RMN y de la similitud en las afinidades de unión para los diferentes sustratos de ARNdh e incluso para ADNdh. Por lo tanto, la hipótesis más fuerte que elaboramos a partir de estos resultados es que este dominio debe cumplir con otras funciones diferentes a la de la localización del miARN dentro del precursor.

Una característica distintiva de los dsRBDs en su capacidad para discriminar entre ADNdh y ARNdh; híbridos ADN-ARN o ácidos nucleicos simple hebra^{71,132,150,221}. Sin embargo, en este trabajo se demostró que DCL1-B se une tanto al ARNdh como al ADNdh con afinidad similar. Las estructuras de los complejos dsRBD-ARNdh determinadas hasta el momento, muestran que la diferenciación del ADNdh se puede explicar por interacciones específicas de la proteína con los grupos fosfato del esqueleto del ácido nucleico y con los grupos 2'OH de los grupos ribosa^{76,111,113,139,213}. En particular, los contactos más específicos incluyen a la hélice $\alpha 1$ [un motivo conservado L(Q/N)(E/D)] y el bucle $\beta 1$ - $\beta 2$ (un residuo de His conservado en el centro del bucle)]. La separación entre estas regiones corresponde a las dimensiones de la forma A de la hélice de

los ácidos nucleicos, lo que también ayudaría a la discriminación de sustrato ¹⁰⁹. Se ha reportado recientemente que el dsRBD no canónico de DICER de levaduras une también ADNdh ¹⁴⁴. En este caso, el dominio presenta una inserción en el bucle localizado entre las láminas $\beta 1$ y $\beta 2$ a la vez que, la hélice $\alpha 1$ adquiere una posición diferente, lo que podría explicar la capacidad de unión dual. En el análisis estructural presentado en este trabajo de tesis, DCL1-B tiene una hélice $\alpha 1$ más corta que los dsRBDs canónicos y los residuos que la preceden se encuentran en un régimen de intercambio conformacional intermedio. Estas características podrían relacionarse con la capacidad de unir tanto ADN como ARNdh. La flexibilidad conformacional de la región precedente a la hélice $\alpha 1$, que podría estar involucrada con las interacciones con el ARNdh y con el ADNdh, podría provocar la heterogeneidad en el reconocimiento observada para DCL1-B.

La dinámica del esqueleto peptídico juega una función vital en la unión a sustratos blanco en una amplia variedad de proteínas relacionadas a contextos funcionales muy diferentes ^{185,222,223}. En el caso de DCL1-B, se comprobó que las regiones canónicas de interacción con el sustrato, muestran un aumento en la flexibilidad de del esqueleto peptídico. Esta movilidad le permitiría al dsRBD adoptar múltiples subestados conformacionales que fluctúan en la escala de ms a μ s que le permitirían unirse a ARNdh en conformaciones ligeramente diferentes. Este comportamiento, ha sido reportado para otros dominios de este tipo como el dsRBD de PKR ²¹⁷, DROSHA, el de DGCR8 ²¹⁸ y el de DICER ²¹⁹ de humanos y el de la ARNasa III de *S. cerevisiae* ²²⁴. Sin embargo, al igual de lo que ocurre con el dsRBD de DICER de humanos, DCL1-B presenta una cuarta región adicional de alta movilidad que incluye al bucle $\beta 2$ - $\beta 3$. Esta región, no forma parte de los motivos de interacción canónicos, sino que se encuentra en el lado opuesto de la zona de interacción con el ARNdh. Estos dos dominios, presentan la particularidad de tener funciones de señales de localización nuclear, por lo que

probablemente, este comportamiento dinámico esté relacionado el reconocimiento de otras proteínas.

Por otro lado, se ha demostrado que DCL1 truncada en el segundo dominio de unión a ARN doble hebra de (*dcl1-9*) falla en la localización en los cuerpos D del núcleo, revelando que este dominio estaría involucrado en la localización subcelular de DCL1⁶⁴. De una manera similar, se ha demostrado que el dsRBD no canónico de DICER de levadura es capaz de mediar la localización nuclear de la proteína¹⁴⁴ mientras que el dsRBD de DICER humana, actúa como una señal de localización nuclear (NLS) no clásica en ausencia del dominio helicasa de la proteína¹³¹. En este trabajo de tesis, se demostró que DCL1-B presenta una localización nuclear, mientras que DCL1-A se ubica principalmente en el citoplasma. Esto podría estar relacionado con la función que este dominio lleva a cabo, direccionando a DCL1 hacia el sitio correcto. Un posible mecanismo para llevar a cabo esta función podría ser la interacción con proteínas específicas. En un trabajo publicado en el año 2005, se reportó por primera vez la interacción de los dsRBDs de HYL1 con el dsRBD C-terminal de DCL1²¹⁶, sin embargo dicha interacción no fue observada en los experimentos *in vitro* que se realizaron durante este trabajo. Hasta el momento, se han identificado numerosas proteínas que participan en el procesamiento de miARNs que podrían interactuar potencialmente con DCL1-B⁶⁷ y que podrían estar relacionadas con la función de localización. Otro mecanismo posible por el cual DCL1-B podría llevar a cabo el direccionamiento al núcleo, involucra la interacción con importinas nucleares, como se reportó para el dsRBD de DICER humana¹³¹. En varios trabajos, se ha demostrado que la función de localización nuclear es independiente de la capacidad de unión al sustrato^{130,131}, en este sentido, DCL1-B podría haber evolucionado de manera independiente de la función de unión a ácidos nucleicos para determinar la localización nuclear de DCL1 a través de

interacciones proteína-proteína específicas. Para lograr demostrar este comportamiento, serán necesarios más estudios *in vivo* que permitan definir las funciones de este dominio en el contexto de la proteína completa. La observación que la fusión DCL1-B:GFP presenta localización nuclear nos provee de una herramienta excelente para evaluar los determinantes estructurales que median esta función *in vivo*.

En este trabajo se ha obtenido la estructura en solución de DCL1-B y se ha demostrado que presenta un plegamiento canónico con variaciones que podrían ser funcionalmente relevantes. Se ha demostrado que este dominio une ARNdh sin preferencia por diferentes regiones del sustrato y que además une ADNdh. Por otro lado se observó que presenta un comportamiento dinámico diferente en el bucle $\beta 2$ - $\beta 3$ que podría relacionarse con funciones distintas a la unión de ARNdh por parte de la proteína. En función de la información biológica disponible hasta el momento, el presente estudio sugiere que este dominio podría estar cumpliendo simultáneamente con la unión a sustratos y con funciones relacionadas al direccionamiento de DCL1 al núcleo a partir de interacciones proteína-proteína.

5 CAPÍTULO 2: *dsRBD1 de HYL1*

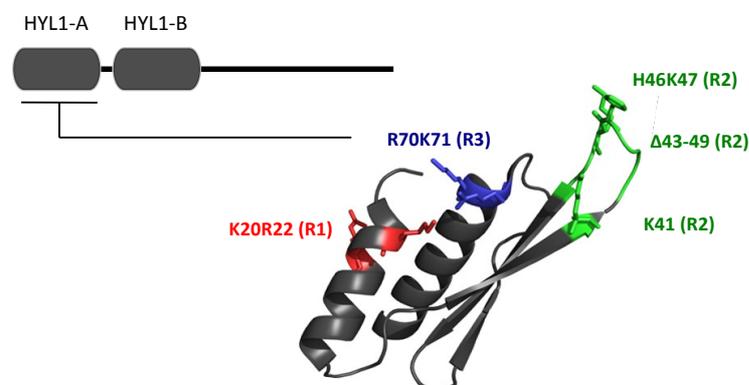
5.1 RESULTADOS 2

5.1.1 RESULTADOS PREVIOS

La proteína HYL1, está involucrada en el procesamiento preciso de los precursores de miRNAs en plantas. Su arquitectura de dominios, incluye dos dsRBDs en el extremos N-terminal, que posiblemente, participen en la selección de sustrato para el procesamiento. En un trabajo previo que se llevó a cabo en nuestro laboratorio, se calcularon las estructuras en solución de ambos dominios dsRBD utilizando restricciones orientacionales sobre la cadena principal, y se determinó la afinidad de los dominios mediante la técnica de retardo en geles con el precursor completo de miR172. De esta forma se pudo observar que HYL1-A posee mayor afinidad por el ARNd_h que HYL1-B. El doble dominio mostró una afinidad similar a la de HYL1-A, indicando que el primer dsRBD sería el principal responsable de la interacción ⁷⁷.

En forma paralela, se llevó a cabo una titulación por RMN de cada dominio con cantidades crecientes de sustrato. En este caso se utilizó el par miR172/miR172* anillado. Esta técnica, además de confirmar los resultados obtenidos en los experimentos de retardo en gel, permitió mapear la superficie de interacción de la proteína con el ARNd_h, mostrando que HYL1-A interacciona *in vitro* con el ARN sustrato a través de las regiones canónicas descritas en las estructuras de complejos ARN:dsRBD ⁷⁷. Una vez identificados los residuos que participan en la unión, se realizaron cuatro mutaciones puntuales y una delección sobre el dominio HYL1-A ubicadas en las regiones de interacción de los dsRBDs con su sustrato (Figura 49). Dichas mutaciones se clonaron en el contexto de la proteína completa en los plásmidos binarios pCHF5 y posteriormente fueron subclonadas en plásmidos de expresión pET-TEV para llevar a cabo una

caracterización biofísica.



Construcción/Planta	Región mutada	Descripción
HYL1	No corresponde	Proteína silvestre
K20R22	Región 1	Mutación puntual K20 y R22 por Ala
K41	Región 2	Mutación puntual K41 por Ala
H46K47	Región 2	Mutación puntual H46 y K47 por Ala
Δ43-49	Región 2	Delección
R70K71	Región 3	Mutación puntual R70 y K71 por Ala

Figura 49: Mapeo de las mutaciones realizadas en las distintas construcciones sobre la estructura de HYL1-A. Abajo, la tabla describe las mutaciones realizadas en las distintas regiones.

5.1.2 FUNCIÓN *IN VIVO* DE LAS VARIANTES DE HYL1-A

Con el objetivo de evaluar si las mutaciones introducidas en el dominio HYL1-A son capaces de afectar la función *in vivo* de la proteína completa, se utilizaron ensayos de complementación de plantas de *Arabidopsis thaliana* con mutaciones nulas en *hyl1* (*hyl1-2*)⁶⁸. Estas plantas, exhiben un fenotipo pleiotrópico caracterizado por hojas hiponásticas, un tamaño de hojas reducidas, disminución del tamaño de la planta, retardo en la floración, fertilidad reducida y múltiples brotes laterales⁷¹. En las plantas *hyl1-2* el procesamiento de miARNs está obstaculizado, llevando a una disminución del producto final, miARNs, y un incremento de los niveles de precursores. Como punto de partida, se

transformaron 20 plantas con cada construcción por el método de “*floral dip*”. Las construcciones incluían al el gen *hyl1* completo con cada una de las cinco mutaciones diseñadas previamente sobre el dsRBD1 (Figura 49). Como control positivo se utilizó el gen *hyl1* silvestre. Todas las secuencias fueron clonadas bajo el control del promotor 35S del virus del mosaico del coliflor (CaMV-35S) y contienen una etiqueta de hemaglutinina (HA). Como control negativo se utilizó el vector pCHF5 vacío. Luego de seleccionar la primera generación (T1) con BASTA, se obtuvieron entre 25 y 50 plantas transgénicas de cada línea. Las mismas, mostraron un fenotipo variable que abarcaba desde un fenotipo *hyl-1* fuerte hasta un fenotipo completamente rescatado, idéntico al de las plantas silvestres. Las plantas se clasificaron a los 20 días de la germinación como: completamente rescatadas (WT), parcialmente rescatadas (*hyl-*) y no rescatadas (*hyl*), en función de las características de las hojas de la roseta (Figura 50).

La clasificación mostró que cada región tenía una tendencia del grado de rescate fenotípico. Inesperadamente, la mayor deficiencia en la función de HYL1 se observó para las mutantes en las regiones R1 y R3, mientras que las mutaciones en la región R2, que en la literatura se propone como una de las principales determinantes de la unión, mostró altos niveles de rescate fenotípico, indicando que no sería de gran importancia para la función de HYL1 *in vivo*. Este último resultado fue llamativo, debido a que muchos estudios previos en proteínas homólogas sugieren que este bucle es esencial para el reconocimiento de ARN por los dsRBDs.

La función *in vivo* de una proteína con una baja actividad en principio puede ser rescatada por altos niveles de expresión que podrían compensar la actividad específica²²⁵. Al considerar que las proteínas fueron expresadas bajo el control del promotor fuerte CaMV35S, se investigó si los fenotipos observados podrían ser el resultado de las diferencias en los niveles de expresión de las

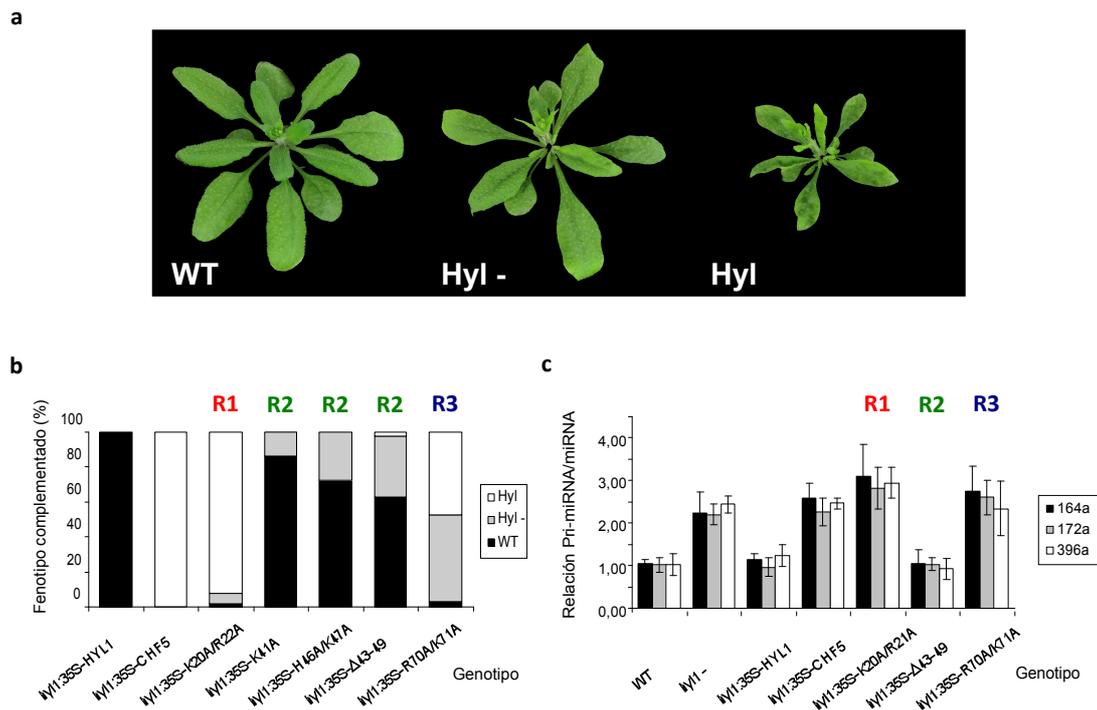


Figura 50: a) Clasificación de fenotipos de las plantas complementadas. Dentro de cada grupo de plantas transgénicas, las plantas se clasificaron como totalmente complementadas (WT), parcialmente complementadas (hyl-) y no complementadas (hyl). b) Análisis fenotípico de las plantas complementadas. Las barras indican el porcentaje de plantas complementadas que fueron clasificadas dentro de cada grupo en función del fenotipo. c) Análisis funcional de las plantas complementadas. Relación de precursores con los miARNs en plantas complementadas con proteínas mutantes para tres precursores diferentes: pri-miR164a, pri-miR172a, pri-miR396a

proteínas. Con el propósito de comprobar esta hipótesis, se determinaron de los niveles de expresión en plantas individuales de las distintas líneas mutantes mediante un análisis de western blot aprovechando la etiqueta de HA que tienen las construcciones (Figura 51). La comparación de los niveles de expresión con el rescate fenotípico de 36 plantas seleccionadas al azar, muestra que no existe correlación entre estas dos variables, lo que estaría demostrando por un lado, que la función de HYL1 no sería dependiente de la cantidad de proteína expresada y, por otro que las variaciones en la complementación del fenotipo observado para las distintas proteínas mutantes estaría reflejando la actividad de

la proteína *in vivo* (Figura 52).

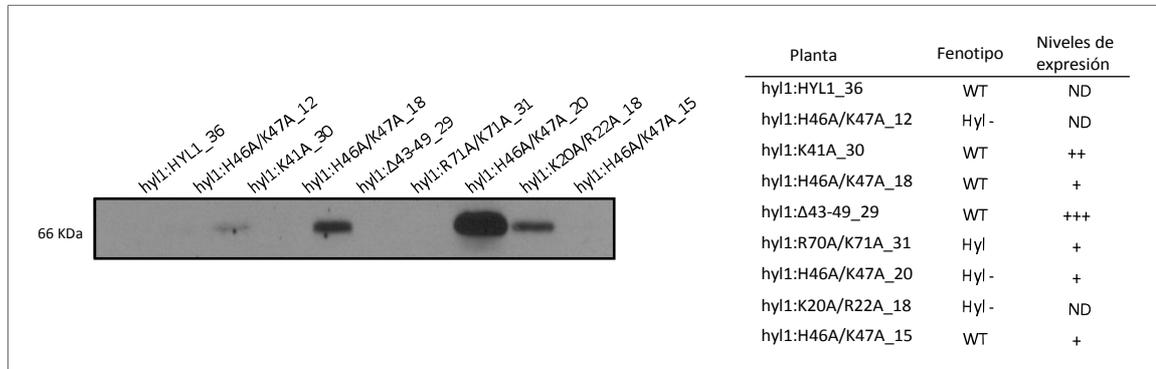


Figura 51: Cuantificación de los niveles de proteína en las distintas plantas transgénicas, a la derecha se muestran los niveles de expresión y la clasificación fenotípica de cada planta.

Actualmente, está bien establecido una de las las funciones de HYL1 es asistir a DCL1 en el procesamiento de miARNs^{63,70}. Con el objetivo de evaluar la función de las proteínas mutantes en una forma más directa, se midieron los niveles de miARNs maduros y de sus precursores en las plantas transgénicas. Para esto, se cuantificaron los pri-miR164, pri-miR172 y pri-miR396 y los respectivos miARNs en plantas silvestres, *hyl1-2* y *hyl1-2* complementadas que expresaban cantidades similares de proteína

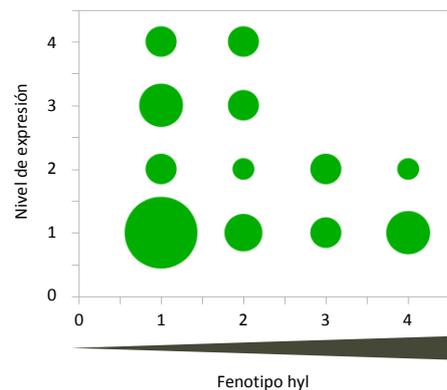


Figura 52: Correlación entre los niveles de expresión y el fenotipo de las plantas complementadas. Al fenotipo silvestre se le asignó un valor de 1 y a el de *hyl* una valor de 4. Los valores 2 y 3 corresponden a fenotipos intermedios. Se asignaron valores de 1 a 4 para los niveles de expresión de proteína en función de la intensidad de las bandas del western blot. El diámetro de los círculos es proporcional a la cantidad de plantas.

(Figura 53). En todos los casos, la actividad de procesamiento se corresponde bien con el rescate fenotípico (Figura 50). A partir de estos resultados, es posible concluir que las mutaciones introducidas afectan la función de HYL-1 al nivel de el procesamiento de pri-miARNs.

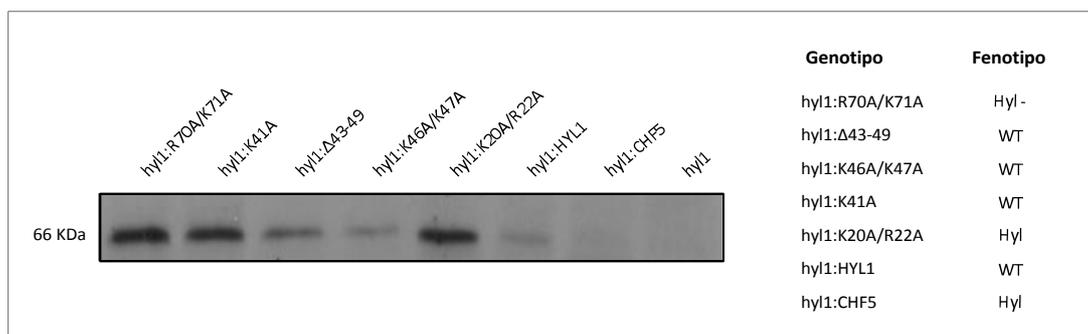


Figura 53: Niveles de expresión de plantas seleccionadas para realizar las cuantificaciones de miARNs y de sus precursores.

5.1.3 ANÁLISIS ESTRUCTURAL DE LOS DOMINIOS MUTANTES

La introducción de mutaciones puede llevar a cambios locales o globales en las estructuras de las proteínas. Con el objetivo de descartar rearrreglos estructurales significativos que pudieran invalidar las conclusiones respecto de la participación de distintas regiones en la función de HYL1-A *in vivo*, se generaron construcciones para la sobreexpresión en *E. coli* de las diferentes variantes de HYL1-A y se marcaron las proteínas con ^{15}N para obtener los espectros $^1\text{H}^{15}\text{N}$ -HSQC para cada una de las mutantes. Este tipo de espectros es muy sensible a alteraciones en las estructuras de proteínas, y permite la localización de las modificaciones con definición al nivel de residuos. La superposición de los espectros del dominio silvestre con los de las distintas mutantes muestra que las alteraciones estructurales están limitadas a las regiones cercanas a la mutación introducida (Figura 54). Esto es así, aún para la mutante que contiene la delección, donde se podrían esperar rearrreglos importantes en la estructura global de la

proteína.

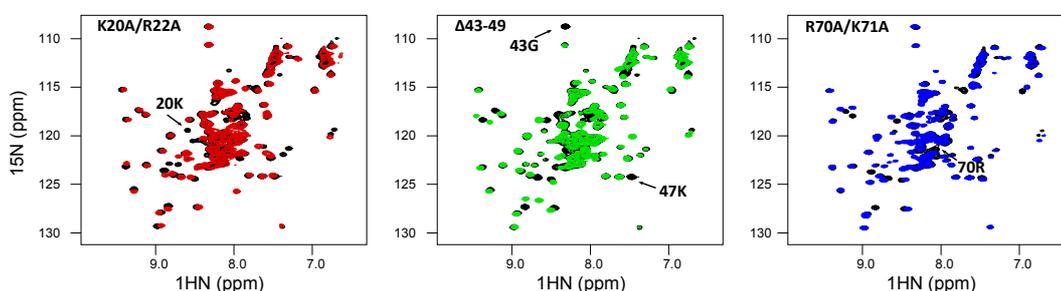


Figura 54: Espectros $^1\text{H}^{15}\text{N}$ -HSQC de las proteínas mutadas comparados con el espectro de la proteína silvestre (negro). Las señales correspondientes a los residuos mutados se destacan con flechas en los espectros de la proteína silvestre.

Con el objetivo de encontrar las bases estructurales de las alteraciones en la función de HYL1 causadas por las mutaciones, se obtuvieron modelos estructurales de las proteínas mutantes basados en la estructura cristalina de la proteína silvestre (3ADG.pdb) usando el programa Rosetta. Como se esperaba, las mutaciones puntuales no llevaron a rearrreglos estructurales grandes, y la delección del bucle β 1- β 2 es bien tolerada para el plegamiento de la proteína (Figura 55).

En las interacciones entre proteínas y ácidos nucleicos las componentes electrostáticas tienen en general una participación preponderante. Para evaluar las alteraciones en el campo eléctrico

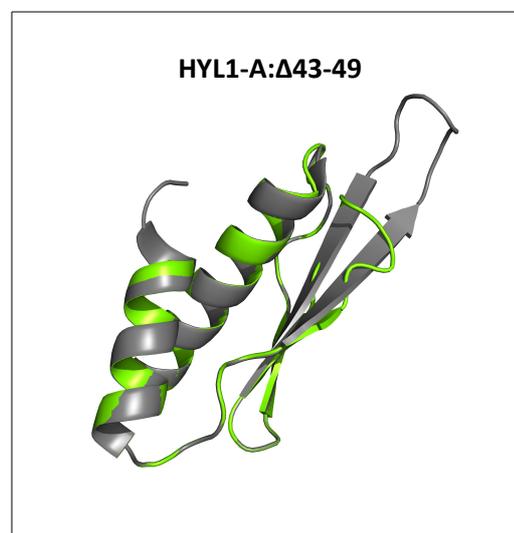


Figura 55: Superposición de la estructura de la mutante Δ 43-49 (R2) modelada con Rosetta con la estructura cristalina de HYL1-A (3ADG.pdb).

alrededor de la proteína introducidas por las mutaciones realizamos un cálculo del campo sobre los modelos de las mutantes utilizando el programa APBS. Al analizar las superficies correspondientes a un potencial fijo en las distintas mutantes, se observó que la diferencia más importante en las variantes, en general, reside en un parche con potencial electrostático positivo formado por un grupo de residuos básicos que se ubican alrededor de las regiones R1 y R3 tal como se observó para el caso de DCL1-B. Estos residuos son los que contactan el esqueleto de fosfatos del ARN a través de interacciones electrostáticas. La eliminación de estas cadenas laterales básicas en la región R1 y R3, llevan a una gran reducción en el potencial positivo. En contraste con esto, las mutaciones en la región R2 muestran un potencial electrostático conservado respecto al de la proteína silvestre (Figura 56). Es decir, que la atracción electrostática hacia el sustrato que llevan a cabo las cadenas laterales de los residuos de las regiones R1 y R3 de HYL1 son esenciales para la función correcta de la proteína.

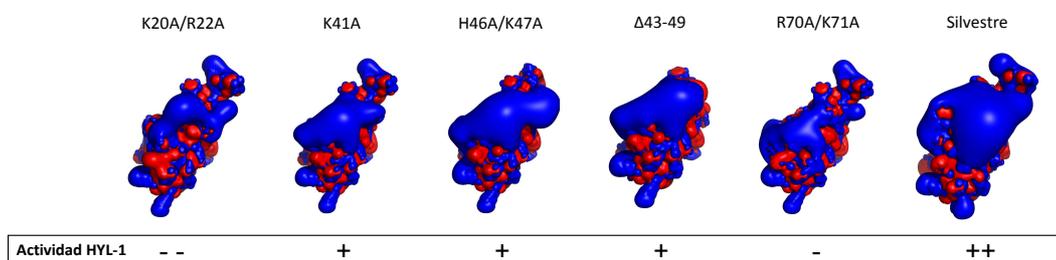


Figura 56: Cálculo del campo electrostático sobre las proteínas silvestre y mutantes. El campo electrostático alrededor de cada proteína fue calculado utilizando APBS²⁰⁹. Se muestran las isosuperficies a 1,5V para el potencial positivo (azul) y -1,5 V para el negativo (rojo).

5.1.4 ESTABILIDAD Y UNIÓN AL SUSTRATO

Una vez que se estableció que todas las mutantes presentaban un plegamiento correcto, se investigó si las diferencias observadas en las actividades

de HYL1 *in vivo* se debían a la desestabilización de la proteína por causa de las mutaciones introducidas. Por este motivo, se determinó la estabilidad (ΔG° N-D) del primer dsRBD de HYL1 y de las mutantes correspondientes utilizando la técnica de dicroísmo circular para monitorear la desnaturalización inducida por urea. La estabilidad de la proteína se encuentra levemente incrementada en las mutaciones de las regiones R2 y R3 respecto a la proteína silvestre, mientras que la estabilidad de la mutante en la región R1 es comparable con la de la proteína silvestre (Figura 57). A partir de estos resultados se puede afirmar que las mutaciones no afectan negativamente la estabilidad de HYL1-A.

Posteriormente, se realizó un ensayo de polarización de fluorescencia para evaluar los efectos de las mutaciones sobre la unión al sustrato. Con este propósito, se utilizaron las construcciones de HYL1-A y sus variantes, y como sustrato, la región correspondiente al tallo inferior del pre-miR172. Todas las proteínas mutantes mostraron una disminución en la afinidad de unión respecto al dominio silvestre. Notablemente, la doble mutación en la región R1 (K20A/R22A) apenas se une al ARN, aún a las mayores concentraciones de proteína utilizadas. La mutación doble localizada en la región R3 (R70A/K71A) también presenta una afinidad más reducida respecto a la proteína silvestre (Figura 57).

Las constantes de unión para estas dos mutantes no pudieron ser estimadas de manera precisa por la falta de puntos correspondientes a valores de anisotropía saturantes debido a que la interferencia de la dispersión de la luz con las medidas de anisotropía obstaculiza la adquisición de puntos a mayores concentraciones de proteína. De manera sorprendente, considerando la importancia adjudicada en trabajos previos a la región R2 de los dsRBDs en la unión, las tres mutaciones en esta región arrojaron una disminución leve en las afinidades de interacción. Las constantes de disociación de estas mutantes son

similares, y se encuentran en el rango de 2,5 a 5,4 μM , dando una visión de conjunto coherente de la influencia de esta región en la afinidad de HYL-1 por el ARN. Aún cuando se elimina el bucle completo, las proteínas conservan la afinidad por el sustrato.

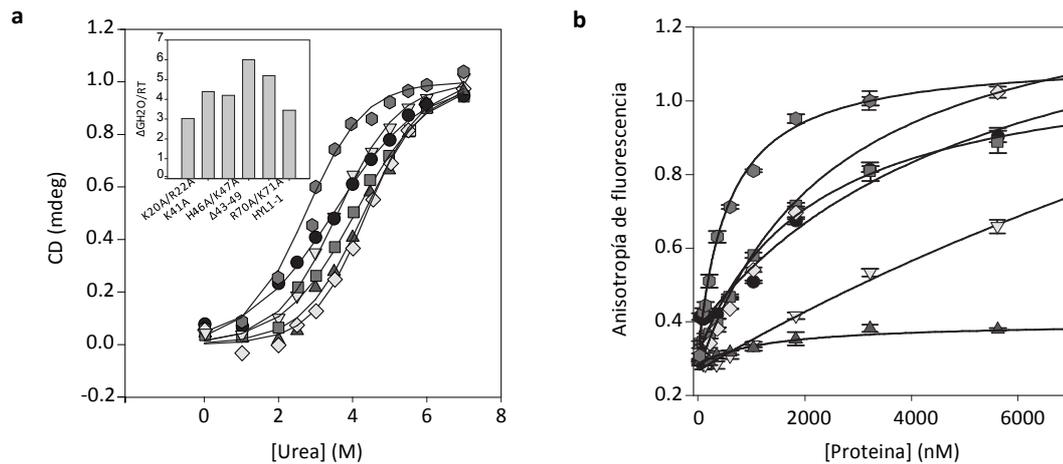


Figura 57: Caracterización biofísica de los dominios dsRBD1 de HYL-1 mutantes. a) Desnaturalización inducida por urea de las proteínas mutantes y del dominio silvestre seguido por dicroísmo circular medido a 220 nm. El recuadro de arriba a la izquierda muestra el $\Delta G^{\circ} H_2O$ (Kcal M^{-1}) para cada una de las proteínas. b) Afinidad por unión al sustrato de los dominios mutantes y silvestre. Se determinó la anisotropía de fluorescencia de pre-miR172ls marcado con fluoresceína con cantidades crecientes de proteína. En a y b los símbolos corresponden a: hexágonos grises, silvestre; triángulos negros, K20R22; rombos grises, K41; cuadrados negros, H46K47; círculos grises, $\Delta 43-49$ y triángulos grises, R70K71.

5.2 DISCUSIÓN 2

La función exacta de HYL1 dentro del complejo de procesamiento de miARN de plantas es hasta el momento desconocida. Se ha demostrado tanto *in vitro* como *in vivo* que la proteína es importante para que el corte del precursor se lleve a cabo de una manera precisa^{63,70}. Trabajos recientes han demostrado que la falta de HYL1 puede ser compensada por una forma más activa de DCL1^{225,226}. En estos casos es posible generar cantidades suficientes de miARNs correctamente procesados, lo que limita los efectos de la imprecisión en el corte. Se ha propuesto que HYL1 también participa en la selección de hebra y distribución de los productos hacia AGO1 dentro del complejo RISC, al igual que lo hace TRBP en humanos⁷⁸. A pesar de que en varios trabajos se han intentado establecer las funciones de las proteínas auxiliares que contienen solamente dominios dsRBDs en otros organismos, hasta el momento, no se ha llegado a un consenso.

Aunque el mecanismo de acción de HYL1 en el procesamiento de miARNs en plantas no está completamente comprendido, la unión al ARN sustrato parece ser la principal función de esta proteína. La unión a sustrato por parte de HYL1 está dominada por el primer dsRBD (HYL1-A), con una contribución marginal de la afinidad de unión por HYL1-B y el resto de la proteína^{76,77}. En este trabajo de tesis se ha diseccionado la función de cada determinante de unión a ARNdh del dominio HYL1-A.

La caracterización que se llevó a cabo en este trabajo mostró que las regiones R1 y R3 del dominio son los determinantes estructurales más importantes en la unión al ARN. En la mayoría de los dsRBDs, la región R3 contiene un motivo KKxAK bien conservado que reconoce el esqueleto fosfodiéster del surco mayor de ARNdh. Estos residuos, forman un parche

electrostático significativo sobre la superficie del dominio creado por las cadenas de lisina cargadas positivamente (Figura 56). En HYL1-A, este motivo es bipartito y se encuentra dividido entre la región R1 y la región R3, donde la tercer Lys del motivo conservado se encuentra reemplazada por un Glu. Sin embargo, esta sustitución es compensada por la presencia de un residuo de Lys en la posición 20 que provee la carga positiva necesaria para interactuar con el esqueleto del ARNdh, y fue mutado por Ala en el presente trabajo. En este caso, el grupo amino terminal de la cadena lateral se ubica en una posición equivalente a la de la Lys de la región R3. Las mutaciones introducidas en las regiones R1 y R3 en el presente trabajo de tesis han llevado a resultados similares respecto a las características estructurales que involucran la irrupción del parche electrostático. En este sentido, se puede racionalizar el impacto similar de ambas mutaciones en la función y en la afinidad de HYL1 por el ARNdh. La ausencia de alteraciones en la estabilidad en las proteínas mutantes en R1 y R3 destaca que el reconocimiento del esqueleto de los fosfatos es esencial para la unión al ARN por parte de estos dominios.

Un resultado inesperado en este trabajo fue que el bucle $\beta 1$ - $\beta 2$ de HYL1-A no es esencial para la actividad de HYL1 y tiene muy poca influencia en la afinidad de unión al ARN. Este resultado es difícil de interpretar desde un punto de vista estructural ya que el bucle juega en principio un papel importante en la unión al ARN insertándose en el surco menor. Dicha interacción se da de manera directa con grupos de la ribosa y con las bases, y contribuye al reconocimiento de la forma A de la hélice de ARN doble hebra, localizándose exactamente a dos vueltas de hélice de la posición de interacción de la región R1. La importancia de este bucle en otros dsRBDs ha sido demostrada a partir de análisis mutacionales¹³⁶. De hecho, este bucle es la región que muestra las mayores diferencias de desplazamientos químicos ante la formación del complejo con ARN⁷⁷. Sin

embargo, en HYL1, el residuo de histidina altamente conservado en la parte superior del bucle puede ser mutado y aún se puede eliminar el bucle completo sin cambios significativos en la afinidad por el ARN, estabilidad de la proteína o función de la proteína. Esto muestra que la importancia del bucle en el reconocimiento del ARN es diferente entre los dsRBDs. En este sentido, es notable que los dsRBDs de las proteínas DICER tienen un bucle más corto que en principio no podría participar en la unión al ARN, al menos en los casos en que la unión sea canónica^{227,228}. Se ha sugerido que la ausencia de este bucle podría dificultar la unión a ARN por estos dominios²²⁸, pero recientemente se observó que el dsRBD de DICER es capaz de unir ARNdh y precursores de miARN²¹⁹. Esta evidencia experimental va de la mano con los resultados obtenidos en esta tesis, mostrando que el bucle β 1- β 2 no es necesario para la unión al ARNdh por HYL1. Para comprender el modo en que estas proteínas llevan a cabo la unión con el sustrato de manera diferencial respecto a los dsRBDs canónicos, será necesario realizar un estudio estructural del complejo formado por estos dominios con bucle corto o por la mutante de HYL1 que incluye la delección.

En un estudio realizado por Manavella y colaboradores, se demostró que HYL1 es regulada por fosforilación⁴². Es notable que uno de los sitios de fosforilación es la serina 39, localizada dentro de este bucle. Cuando este residuo de serina se muta por ácido aspártico, imitando la fosforilación, la función de la proteína es inhibida, mientras que cuando se elimina este sitio de fosforilación por una mutación de serina por alanina, la proteína resultante es completamente funcional. Considerando estos resultados, se podría especular que las mutaciones en la región R2 podrían evitar una posible inhibición de la actividad de HYL1 por fosforilación, por lo que se estaría compensando la pérdida parcial de afinidad introducida por la mutación.

En conclusión, en este trabajo se pudo demostrar que existe una

correlación directa entre la afinidad de unión a sustrato por HYL1-A y su actividad *in vivo*. Por otro lado, también se ha establecido que la unión de HYL1-A al ARNdh es esencial para su función y que los determinantes estructurales de unión que corresponden a las regiones R1 y R3, son necesarios para que la proteína lleve a cabo su función *in vivo* de manera correcta. Finalmente, el análisis de líneas que sobreexpresan las proteínas demuestra que su actividad no puede ser compensada por exceso de proteína.

6 CONCLUSIONES GENERALES

En el presente trabajo de tesis se han estudiado y caracterizado biofísica y funcionalmente los dsRBDs de las proteínas de procesamiento de miARNs de plantas, DCL1 y HYL1.

En el caso de **DCL1**, se determinó la estructura en solución de DCL1-B por RMN, a partir de la cual se pudo observar que si bien este dominio presenta la topología canónica de dsRBD, presenta algunas variaciones en el plegamiento que lo distinguen del resto. DCL1-B presenta una hélice $\alpha 1$ más corta, lo que podría estar directamente relacionado con las capacidades de unión al sustrato y además tiene una inserción en el bucle $\beta 2$ - $\beta 3$, ubicado en el lado opuesto de la superficie de interacción con los ácidos nucleicos. Se observó que dicha inserción presenta un residuo de Trp expuesto al solvente que podría llevar a cabo funciones de interacción con otros dominios proteicos. Si bien está reportado que este dominio es capaz de interactuar con los dsRBDs de HYL1, no se observó dicha interacción de manera directa mediante experimentos *in vitro*. También se estableció que DCL1-B es incapaz de llevar a cabo un reconocimiento específico de ácidos nucleicos. En este trabajo se evaluó la capacidad de unión de este dominio con diferentes sustratos de ARNdh y ADNdh. Se observó que DCL1-B no muestra diferencias significativas en la afinidad de unión entre ARNdh perfectamente apareado, ARNdh con imperfecciones y ADNdh. Al caracterizar la interacción de DCL1-B con los sustratos ARN y ADN doble hebra por Resonancia Magnética Nuclear, se observó que este dominio forma complejos heterogéneos con los ácidos nucleicos y que presenta un régimen de intercambio intermedio. Las regiones de interacción mapeadas en la estructura se localizan en R1, R2 y R3 al igual que en los dsRBDs canónicos. Respecto a la dinámica de este dominio, se observó que las regiones que presentan mayor movilidad son aquellas que

participarían de la unión a ARNdh pero que también existe una región flexible en la cara opuesta de la zona de interacción con ácidos nucleicos localizada en el bucle $\beta 2$ - $\beta 3$. Respecto a la caracterización funcional de este dominio, se estableció que a diferencia de DCL1-A, DCL1-B es capaz de mediar la localización nuclear de la proteína reportera GFP de *Aequorea victoria* en células radiculares de *Arabidopsis thaliana*.

En el caso de **HYL1** se lograron identificar los determinantes de unión a ARNdh que son esenciales para que la proteína lleve a cabo una correcta función *in vivo*. Se estableció que las regiones de interacción R1 y R3 de HYL1-A son las que limitan la actividad de la proteína, afectando el reconocimiento del surco mayor de la forma A del ARNdh. Mientras que el reconocimiento del surco menor por parte de R2 no sería limitante de la actividad *in vivo*. Por otro lado se determinó que la deficiencia en la actividad de las proteínas con mutaciones en R1 y R3 de HYL1-A estarían afectando la síntesis de miARNs a nivel del procesamiento, debido a que estas plantas presentan altas relaciones de precursores:miARNs, cosa que no ocurre en plantas deficientes en la región R2 de este dominio. A su vez se observó que la actividad de HYL1 deficiente no puede ser suplantada por un aumento en los niveles de expresión de proteína recombinante, reflejando la importancia de la estructura en estas regiones del dominio para que HYL1 pueda llevar a cabo su función. También se observó que la introducción de mutaciones en las regiones canónicas de interacción de HYL1-A no introducen cambios en el plegamiento global del dominio, ni siquiera la deleción del bucle $\beta 1$ - $\beta 2$ completo. Tampoco se afecta de manera negativa la estabilidad de las proteínas. Un punto importante que se pudo establecer mediante este trabajo fue que la modificación de las regiones R1 y R3 del primer dsRBD de HYL1 afectan directamente la afinidad por el sustrato ARNdh, lo que lleva a una disminución del procesamiento y por lo tanto, está vinculado

directamente con la actividad de la proteína *in vivo*.

7 BIBLIOGRAFÍA

1. Napoli, C., Lemieux, C. & Jorgensen, R. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell* **2**, 279–289 (1990).
2. Romano, N. & Macino, G. Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Mol. Microbiol.* **6**, 3343–53 (1992).
3. Guo, S. & Kemphues, K. J. *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81**, 611–20 (1995).
4. Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950–2 (1999).
5. Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293–6 (2000).
6. Zamore, P. D., Tuschl, T., Sharp, P. A. & Bartel, D. P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**, 25–33 (2000).
7. Chapman, E. J. & Carrington, J. C. Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.* **8**, 884–96 (2007).
8. Qi, Y., Denli, A. M. & Hannon, G. J. Biochemical specialization within Arabidopsis RNA silencing pathways. *Mol. Cell* **19**, 421–8 (2005).
9. Vaucheret, H. Plant ARGONAUTES. *Trends Plant Sci.* **13**, 350–8 (2008).
10. Ramachandran, V. & Chen, X. Small RNA metabolism in Arabidopsis. *Trends Plant Sci.* **13**, 368–74 (2008).
11. Vaucheret, H. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev.* **20**, 759–71 (2006).
12. Voinnet, O. Origin, biogenesis, and activity of plant microRNAs. *Cell* **136**, 669–87 (2009).
13. Mello, C. C. & Conte, D. Revealing the world of RNA interference. *Nature* **431**, 338–42 (2004).
14. Xie, Z. & Qi, X. Diverse small RNA-directed silencing pathways in plants. *Biochim. Biophys. Acta* **1779**, 720–4 (2008).
15. Allen, E. & Howell, M. D. miRNAs in the biogenesis of trans-acting siRNAs in higher plants. *Semin. Cell Dev. Biol.* **21**, 798–804 (2010).

16. Allen, E., Xie, Z., Gustafson, A. M. & Carrington, J. C. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**, 207–21 (2005).
17. Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H. L. & Poethig, R. S. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes Dev.* **18**, 2368–79 (2004).
18. Borsani, O., Zhu, J., Verslues, P. E., Sunkar, R. & Zhu, J.-K. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* **123**, 1279–91 (2005).
19. Katiyar-Agarwal, S. *et al.* A pathogen-inducible endogenous siRNA in plant immunity. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 18002–7 (2006).
20. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
21. Llave, C., Xie, Z., Kasschau, K. D. & Carrington, J. C. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* **297**, 2053–6 (2002).
22. Aukerman, M. J. & Sakai, H. Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell* **15**, 2730–41 (2003).
23. Chen, X. A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. *Science* **303**, 2022–5 (2004).
24. Kim, V. N. Minireview Molecules and Small RNAs : Classification , Biogenesis , and Function. **19**, 1–15 (2005).
25. Kim, V. N. & Nam, J.-W. Genomics of microRNA. *Trends Genet.* **22**, 165–73 (2006).
26. Xie, Z. *et al.* Expression of Arabidopsis MIRNA genes. *Plant Physiol.* **138**, 2145–54 (2005).
27. Gregory, R. I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235–40 (2004).
28. Han, J. *et al.* Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**, 887–901 (2006).
29. Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **10**, 126–39 (2009).
30. Förstemann, K. *et al.* Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* **3**, e236 (2005).
31. Saito, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. *PLoS Biol.* **3**, e235 (2005).

32. Chendrimada, T. P. *et al.* TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740–4 (2005).
33. Haase, A. D. *et al.* TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep.* **6**, 961–7 (2005).
34. Lee, Y. *et al.* The role of PACT in the RNA silencing pathway. *EMBO J.* **25**, 522–32 (2006).
35. Kok, K. H., Ng, M.-H. J., Ching, Y.-P. & Jin, D.-Y. Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA. *J. Biol. Chem.* **282**, 17649–57 (2007).
36. Park, W., Li, J., Song, R., Messing, J. & Chen, X. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12**, 1484–95 (2002).
37. Kurihara, Y. & Watanabe, Y. *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12753–8 (2004).
38. Lobbes, D., Rallapalli, G., Schmidt, D. D., Martin, C. & Clarke, J. SERRATE: a new player on the plant microRNA scene. *EMBO Rep.* **7**, 1052–8 (2006).
39. Yang, L., Liu, Z., Lu, F., Dong, A. & Huang, H. SERRATE is a novel nuclear regulator in primary microRNA processing in *Arabidopsis*. *Plant J.* **47**, 841–50 (2006).
40. Gregory, B. D. *et al.* A link between RNA metabolism and silencing affecting *Arabidopsis* development. *Dev. Cell* **14**, 854–66 (2008).
41. Laubinger, S. *et al.* Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8795–800 (2008).
42. Manavella, P. A. *et al.* Fast-forward genetics identifies plant CPL phosphatases as regulators of miRNA processing factor HYL1. *Cell* **151**, 859–70 (2012).
43. Yu, B. *et al.* The FHA domain proteins DAWDLE in *Arabidopsis* and SNIP1 in humans act in small RNA biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 10073–8 (2008).
44. Ren, G. *et al.* Regulation of miRNA abundance by RNA binding protein TOUGH in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 12817–21 (2012).
45. Zhan, X. *et al.* *Arabidopsis* proline-rich protein important for development and abiotic stress tolerance is involved in microRNA biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18198–203 (2012).
46. Wu, X. *et al.* A role for the RNA-binding protein MOS2 in microRNA maturation in *Arabidopsis*. *Cell Res.* **23**, 645–57 (2013).
47. Boutet, S. *et al.* *Arabidopsis* HEN1: a genetic link between endogenous miRNA controlling

- development and siRNA controlling transgene silencing and virus resistance. *Curr. Biol.* **13**, 843–8 (2003).
48. Park, M. Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H. & Poethig, R. S. Nuclear processing and export of microRNAs in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 3691–6 (2005).
49. Doyle, M., Jaskiewicz, L. & Filipowicz, W. in *Enzym.* **32**, 1–35 (2012).
50. Macrae, I. J. *et al.* Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195–8 (2006).
51. Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E. & Filipowicz, W. Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**, 57–68 (2004).
52. MacRae, I. J., Zhou, K. & Doudna, J. A. Structural determinants of RNA recognition and cleavage by Dicer. *Nat. Struct. Mol. Biol.* **14**, 934–40 (2007).
53. Park, J.-E. *et al.* Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* **475**, 201–5 (2011).
54. Ma, E., MacRae, I. J., Kirsch, J. F. & Doudna, J. A. Autoinhibition of human dicer by its internal helicase domain. *J. Mol. Biol.* **380**, 237–43 (2008).
55. Ma, E., Zhou, K., Kidwell, M. A. & Doudna, J. A. Coordinated activities of human dicer domains in regulatory RNA processing. *J. Mol. Biol.* **422**, 466–76 (2012).
56. Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H. & Tomari, Y. Recognition of the pre-miRNA structure by Drosophila Dicer-1. *Nat. Struct. Mol. Biol.* **18**, 1153–8 (2011).
57. Lau, P.-W., Potter, C. S., Carragher, B. & MacRae, I. J. Structure of the human Dicer-TRBP complex by electron microscopy. *Structure* **17**, 1326–32 (2009).
58. Chakravarthy, S., Sternberg, S. H., Kellenberger, C. A. & Doudna, J. A. Substrate-specific kinetics of Dicer-catalyzed RNA processing. *J. Mol. Biol.* **404**, 392–402 (2010).
59. Jinek, M. & Doudna, J. A. A three-dimensional view of the molecular machinery of RNA interference. *Nature* **457**, 405–12 (2009).
60. Lau, P.-W. *et al.* The molecular architecture of human Dicer. *Nat. Struct. Mol. Biol.* **19**, 436–40 (2012).
61. Schauer, S. E., Jacobsen, S. E., Meinke, D. W. & Ray, A. DICER-LIKE1: blind men and elephants in Arabidopsis development. *Trends Plant Sci.* **7**, 487–91 (2002).
62. Nodine, M. D. & Bartel, D. P. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes Dev.* **24**, 2678–92 (2010).
63. Dong, Z., Han, M.-H. & Fedoroff, N. The RNA-binding proteins HYL1 and SE promote accurate

- in vitro processing of pri-miRNA by DCL1. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9970–5 (2008).
64. Fang, Y. & Spector, D. L. Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants. *Curr. Biol.* **17**, 818–23 (2007).
65. Wassenegger, M. The role of the RNAi machinery in heterochromatin formation. *Cell* **122**, 13–6 (2005).
66. Bologna, N. G., Mateos, J. L., Bresso, E. G. & Palatnik, J. F. A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *EMBO J.* **28**, 3646–56 (2009).
67. Xie, Z. Piecing the puzzle together: genetic requirements for miRNA biogenesis in Arabidopsis thaliana. *Methods Mol. Biol.* **592**, 1–17 (2010).
68. Song, L., Han, M.-H., Lesicka, J. & Fedoroff, N. Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 5437–42 (2007).
69. Fujioka, Y., Utsumi, M., Ohba, Y. & Watanabe, Y. Location of a possible miRNA processing site in SmD3/SmB nuclear bodies in Arabidopsis. *Plant Cell Physiol.* **48**, 1243–53 (2007).
70. Kurihara, Y., Takashi, Y. & Watanabe, Y. The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA* **12**, 206–12 (2006).
71. Lu, C. & Fedoroff, N. A mutation in the Arabidopsis HYL1 gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. *Plant Cell* **12**, 2351–2366 (2000).
72. Han, M.-H., Goud, S., Song, L. & Fedoroff, N. The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1093–8 (2004).
73. Vazquez, F., Gascioli, V., Crété, P. & Vaucheret, H. The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Curr. Biol.* **14**, 346–51 (2004).
74. Grigg, S. P., Canales, C., Hay, A. & Tsiantis, M. SERRATE coordinates shoot meristem function and leaf axial patterning in Arabidopsis. *Nature* **437**, 1022–6 (2005).
75. Wu, F. *et al.* The N-terminal double-stranded RNA binding domains of Arabidopsis HYPONASTIC LEAVES1 are sufficient for pre-microRNA processing. *Plant Cell* **19**, 914–25 (2007).
76. Yang, S. W. *et al.* Structure of Arabidopsis HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. *Structure* **18**, 594–605 (2010).

77. Rasia, R. M. *et al.* Structure and RNA interactions of the plant MicroRNA processing-associated protein HYL1. *Biochemistry* **49**, 8237–9 (2010).
78. Eamens, A. L., Smith, N. A., Curtin, S. J., Wang, M.-B. & Waterhouse, P. M. The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA* **15**, 2219–35 (2009).
79. Schwartz, B. W., Yeung, E. C. & Meinke, D. W. Disruption of morphogenesis and transformation of the suspensor in abnormal sus sensor mutants of Arabidopsis. **3245**, 3235–3245 (1994).
80. Prigge, M. J. & Wagner, D. R. The arabidopsis serrate gene encodes a zinc-finger protein required for normal shoot development. *Plant Cell* **13**, 1263–79 (2001).
81. Golling, G. *et al.* Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. *Nat. Genet.* **31**, 135–40 (2002).
82. Machida, S., Chen, H.-Y. & Adam Yuan, Y. Molecular insights into miRNA processing by Arabidopsis thaliana SERRATE. *Nucleic Acids Res.* **39**, 7828–36 (2011).
83. Lewis, J. D. & Izaurralde, E. The role of the cap structure in RNA processing and nuclear export. *Eur. J. Biochem.* **247**, 461–9 (1997).
84. Hugouvieux, V., Kwak, J. M. & Schroeder, J. I. An mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in Arabidopsis. *Cell* **106**, 477–87 (2001).
85. Papp, I., Mur, L. A., Dalmadi, A., Dulai, S. & Koncz, C. A mutation in the Cap Binding Protein 20 gene confers drought tolerance to Arabidopsis. *Plant Mol. Biol.* **55**, 679–86 (2004).
86. Kim, S. *et al.* Two cap-binding proteins CBP20 and CBP80 are involved in processing primary MicroRNAs. *Plant Cell Physiol.* **49**, 1634–44 (2008).
87. Izaurralde, E. *et al.* A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78**, 657–68 (1994).
88. Kawashima, C. G. *et al.* Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. *Plant J.* **57**, 313–21 (2009).
89. Xie, Z., Kasschau, K. D. & Carrington, J. C. Negative feedback regulation of Dicer-Like1 in Arabidopsis by microRNA-guided mRNA degradation. *Curr. Biol.* **13**, 784–9 (2003).
90. Liu, Q., Feng, Y. & Zhu, Z. Dicer-like (DCL) proteins in plants. *Funct. Integr. Genomics* **9**, 277–86 (2009).
91. Rhoades, M. W. *et al.* Prediction of plant microRNA targets. *Cell* **110**, 513–20 (2002).
92. Vaucheret, H., Vazquez, F., Crété, P. & Bartel, D. P. The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev.* **18**, 1187–97 (2004).

93. Mallory, A. C. & Vaucheret, H. ARGONAUTE 1 homeostasis invokes the coordinate action of the microRNA and siRNA pathways. *EMBO Rep.* **10**, 521–6 (2009).
94. Ramachandran, V. & Chen, X. Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. *Science* **321**, 1490–2 (2008).
95. Kennedy, S., Wang, D. & Ruvkun, G. A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* **427**, 645–9 (2004).
96. Chatterjee, S. & Grosshans, H. Active turnover modulates mature microRNA activity in *Caenorhabditis elegans*. *Nature* **461**, 546–9 (2009).
97. Kai, Z. S. & Pasquinelli, A. E. MicroRNA assassins: factors that regulate the disappearance of miRNAs. *Nat. Struct. Mol. Biol.* **17**, 5–10 (2010).
98. Pouch-Pélissier, M.-N. *et al.* SINE RNA induces severe developmental defects in Arabidopsis thaliana and interacts with HYL1 (DRB1), a key member of the DCL1 complex. *PLoS Genet.* **4**, e1000096 (2008).
99. Chu, C.-Y. & Rana, T. M. Small RNAs: regulators and guardians of the genome. *J. Cell. Physiol.* **213**, 412–9 (2007).
100. Cuperus, J. T., Fahlgren, N. & Carrington, J. C. Evolution and functional diversification of MIRNA genes. *Plant Cell* **23**, 431–42 (2011).
101. Mateos, J. L., Bologna, N. G., Chorostecki, U. & Palatnik, J. F. Identification of microRNA processing determinants by random mutagenesis of Arabidopsis MIR172a precursor. *Curr. Biol.* **20**, 49–54 (2010).
102. Song, L., Axtell, M. J. & Fedoroff, N. V. RNA secondary structural determinants of miRNA precursor processing in Arabidopsis. *Curr. Biol.* **20**, 37–41 (2010).
103. Werner, S., Wollmann, H., Schneeberger, K. & Weigel, D. Structure determinants for accurate processing of miR172a in Arabidopsis thaliana. *Curr. Biol.* **20**, 42–8 (2010).
104. Bologna, N. G. *et al.* Multiple RNA recognition patterns during microRNA biogenesis in plants. *Genome Res.* 1675–1689 (2013). doi:10.1101/gr.153387.112
105. Addo-Quaye, C. *et al.* Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the Physcomitrella patens degradome. *RNA* **15**, 2112–21 (2009).
106. Rajagopalan, R., Vaucheret, H., Trejo, J. & Bartel, D. P. A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev.* **20**, 3407–25 (2006).
107. Axtell, M. J. & Bowman, J. L. Evolution of plant microRNAs and their targets. *Trends Plant Sci.* **13**, 343–9 (2008).
108. Vazquez, F. Arabidopsis endogenous small RNAs: highways and byways. *Trends Plant Sci.*

- 11**, 460–8 (2006).
109. Tian, B., Bevilacqua, P. C., Diegelman-Parente, A. & Mathews, M. B. The double-stranded-RNA-binding motif: interference and much more. *Nat. Rev. Mol. Cell Biol.* **5**, 1013–23 (2004).
110. Qin, H. *et al.* Structure of the Arabidopsis thaliana DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. *RNA* **16**, 474–81 (2010).
111. Ryter, J. M. & Schultz, S. C. Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J.* **17**, 7505–13 (1998).
112. Ramos, A., Bayer, P. & Varani, G. Determination of the structure of the RNA complex of a double-stranded RNA-binding domain from Drosophila Staufen protein. *Biopolymers* **52**, 181–96
113. Blaszczyk, J. *et al.* Noncatalytic assembly of ribonuclease III with double-stranded RNA. *Structure* **12**, 457–66 (2004).
114. Wu, H., Henras, A., Chanfreau, G. & Feigon, J. Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8307–12 (2004).
115. Stefl, R., Xu, M., Skrisovska, L., Emeson, R. B. & Allain, F. H.-T. Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. *Structure* **14**, 345–55 (2006).
116. Patel, R. C. & Sen, G. C. PACT, a protein activator of the interferon-induced protein kinase, PKR. *EMBO J.* **17**, 4379–90 (1998).
117. Daher, A. *et al.* Two dimerization domains in the trans-activation response RNA-binding protein (TRBP) individually reverse the protein kinase R inhibition of HIV-1 long terminal repeat expression. *J. Biol. Chem.* **276**, 33899–905 (2001).
118. Peters, G. A., Hartmann, R., Qin, J. & Sen, G. C. Modular structure of PACT: distinct domains for binding and activating PKR. *Mol. Cell. Biol.* **21**, 1908–20 (2001).
119. Romano, P. R. *et al.* Inhibition of double-stranded RNA-dependent protein kinase PKR by vaccinia virus E3: role of complex formation and the E3 N-terminal domain. *Mol. Cell. Biol.* **18**, 7304–16 (1998).
120. Schuldt, A. J. *et al.* Miranda mediates asymmetric protein and RNA localization in the developing nervous system. *Genes Dev.* **12**, 1847–57 (1998).
121. Hung, M.-L., Chao, P. & Chang, K.-Y. dsRBM1 and a proline-rich domain of RNA helicase A can form a composite binder to recognize a specific dsDNA. *Nucleic Acids Res.* **31**, 5741–53 (2003).

122. Tian, B. & Mathews, M. B. Functional characterization of and cooperation between the double-stranded RNA-binding motifs of the protein kinase PKR. *J. Biol. Chem.* **276**, 9936–44 (2001).
123. Hitti, E., Neunteufl, A. & Jantsch, M. F. The double-stranded RNA-binding protein X1rbpa promotes RNA strand annealing. *Nucleic Acids Res.* **26**, 4382–8 (1998).
124. Zheng, X. & Bevilacqua, P. C. Straightening of bulged RNA by the double-stranded RNA-binding domain from the protein kinase PKR. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 14162–7 (2000).
125. Zhang, F. *et al.* Binding of double-stranded RNA to protein kinase PKR is required for dimerization and promotes critical autophosphorylation events in the activation loop. *J. Biol. Chem.* **276**, 24946–58 (2001).
126. Mellits, K. H., Kostura, M. & Mathews, M. B. Interaction of adenovirus VA RNA1 with the protein kinase DA1: nonequivalence of binding and function. *Cell* **61**, 843–52 (1990).
127. Doyle, M. & Jantsch, M. F. Distinct in vivo roles for double-stranded RNA-binding domains of the *Xenopus* RNA-editing enzyme ADAR1 in chromosomal targeting. *J. Cell Biol.* **161**, 309–19 (2003).
128. Eckmann, C. R., Neunteufl, A., Pfaffstetter, L. & Jantsch, M. F. The human but not the *Xenopus* RNA-editing enzyme ADAR1 has an atypical nuclear localization signal and displays the characteristics of a shuttling protein. *Mol. Biol. Cell* **12**, 1911–24 (2001).
129. Strehblow, A., Hallegger, M. & Jantsch, M. F. Nucleocytoplasmic distribution of human RNA-editing enzyme ADAR1 is modulated by double-stranded RNA-binding domains, a leucine-rich export signal, and a putative dimerization domain. *Mol. Biol. Cell* **13**, 3822–35 (2002).
130. Emmerth, S. *et al.* Nuclear retention of fission yeast Dicer is a prerequisite for RNAi-mediated heterochromatin assembly. *Dev. Cell* **18**, 102–13 (2010).
131. Doyle, M. *et al.* The double-stranded RNA binding domain of human Dicer functions as a nuclear localization signal. *RNA* **19**, 1238–52 (2013).
132. St Johnston, D., Brown, N. H., Gall, J. G. & Jantsch, M. A conserved double-stranded RNA-binding domain. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10979–83 (1992).
133. Thomis, D. C. & Samuel, C. E. Mechanism of interferon action: autoregulation of RNA-dependent P1/eIF-2 alpha protein kinase (PKR) expression in transfected mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10837–41 (1992).
134. Feng, G. S., Chong, K., Kumar, A. & Williams, B. R. Identification of double-stranded RNA-binding domains in the interferon-induced double-stranded RNA-activated p68 kinase. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 5447–51 (1992).

135. Green, S. R. & Mathews, M. B. Two RNA-binding motifs in the double-stranded RNA-activated protein kinase, DAI. *Genes Dev.* **6**, 2478–90 (1992).
136. Krovat, B. C. & Jantsch, M. F. Comparative mutational analysis of the double-stranded RNA binding domains of *Xenopus laevis* RNA-binding protein A. *J. Biol. Chem.* **271**, 28112–9 (1996).
137. Kharrat, A., Macias, M. J., Gibson, T. J., Nilges, M. & Pastore, A. Structure of the dsRNA binding domain of *E. coli* RNase III. *EMBO J.* **14**, 3572–84 (1995).
138. Bycroft, M., Grünert, S., Murzin, A. G., Proctor, M. & St Johnston, D. NMR solution structure of a dsRNA binding domain from *Drosophila* staufer protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J.* **14**, 3563–71 (1995).
139. Ramos, A. *et al.* RNA recognition by a Staufer double-stranded RNA-binding domain. *EMBO J.* **19**, 997–1009 (2000).
140. Green, S. R., Manche, L. & Mathews, M. B. Two functionally distinct RNA-binding motifs in the regulatory domain of the protein kinase DAI. *Mol. Cell. Biol.* **15**, 358–64 (1995).
141. McMillan, N. A. *et al.* Mutational analysis of the double-stranded RNA (dsRNA) binding domain of the dsRNA-activated protein kinase, PKR. *J. Biol. Chem.* **270**, 2601–6 (1995).
142. Patel, R. C., Stanton, P. & Sen, G. C. Specific mutations near the amino terminus of double-stranded RNA-dependent protein kinase (PKR) differentially affect its double-stranded RNA binding and dimerization properties. *J. Biol. Chem.* **271**, 25657–63 (1996).
143. Leulliot, N. *et al.* A new alpha-helical extension promotes RNA binding by the dsRBD of Rnt1p RNase III. *EMBO J.* **23**, 2468–77 (2004).
144. Barraud, P. *et al.* An extended dsRBD with a novel zinc-binding motif mediates nuclear retention of fission yeast Dicer. *EMBO J.* **30**, 4223–35 (2011).
145. Huang, Y. *et al.* Structural insights into mechanisms of the small RNA methyltransferase HEN1. *Nature* **461**, 823–7 (2009).
146. Brownawell, A. M. & Macara, I. G. Exportin-5, a novel karyopherin, mediates nuclear export of double-stranded RNA binding proteins. *J. Cell Biol.* **156**, 53–64 (2002).
147. Gwizdek, C. *et al.* Minihelix-containing RNAs mediate exportin-5-dependent nuclear export of the double-stranded RNA-binding protein ILF3. *J. Biol. Chem.* **279**, 884–91 (2004).
148. Macchi, P. *et al.* The brain-specific double-stranded RNA-binding protein Staufer2: nucleolar accumulation and isoform-specific exportin-5-dependent export. *J. Biol. Chem.* **279**, 31440–4 (2004).
149. Fritz, J. *et al.* RNA-regulated interaction of transportin-1 and exportin-5 with the double-stranded RNA-binding domain regulates nucleocytoplasmic shuttling of ADAR1. *Mol. Cell. Biol.* **29**, 1487–97 (2009).

150. Bevilacqua, P. C. & Cech, T. R. Minor-groove recognition of double-stranded RNA by the double-stranded RNA-binding domain from the RNA-activated protein kinase PKR. *Biochemistry* **35**, 9983–94 (1996).
151. Bevilacqua, P. C., George, C. X., Samuel, C. E. & Cech, T. R. Binding of the protein kinase PKR to RNAs with secondary structure defects: role of the tandem A-G mismatch and noncontiguous helices. *Biochemistry* **37**, 6303–16 (1998).
152. Liu, Y., Lei, M. & Samuel, C. E. Chimeric double-stranded RNA-specific adenosine deaminase ADAR1 proteins reveal functional selectivity of double-stranded RNA-binding domains from ADAR1 and protein kinase PKR. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 12541–6 (2000).
153. Nagel, R. & Ares, M. Substrate recognition by a eukaryotic RNase III: the double-stranded RNA-binding domain of Rnt1p selectively binds RNA containing a 5'-AGNN-3' tetraloop. *RNA* **6**, 1142–56 (2000).
154. Doyle, M. & Jantsch, M. F. New and old roles of the double-stranded RNA-binding domain. *J. Struct. Biol.* **140**, 147–53 (2003).
155. Gan, J. *et al.* Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell* **124**, 355–66 (2006).
156. Margis, R. *et al.* The evolution and diversification of Dicers in plants. *FEBS Lett.* **580**, 2442–50 (2006).
157. Han, J. *et al.* The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* **18**, 3016–27 (2004).
158. Liu, Q. *et al.* R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. *Science* **301**, 1921–5 (2003).
159. Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–55 (2009).
160. Liu, X. *et al.* Dicer-1, but not Loquacious, is critical for assembly of miRNA-induced silencing complexes. *RNA* **13**, 2324–9 (2007).
161. Pham, J. W., Pellino, J. L., Lee, Y. S., Carthew, R. W. & Sontheimer, E. J. A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila. *Cell* **117**, 83–94 (2004).
162. Jiang, F. *et al.* Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. *Genes Dev.* **19**, 1674–9 (2005).
163. Zhou, R. *et al.* Processing of Drosophila endo-siRNAs depends on a specific Loquacious isoform. *RNA* **15**, 1886–95 (2009).
164. Hartig, J. V., Esslinger, S., Böttcher, R., Saito, K. & Förstemann, K. Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. *EMBO J.* **28**, 2932–44 (2009).

165. Tomari, Y., Matranga, C., Haley, B., Martinez, N. & Zamore, P. D. A protein sensor for siRNA asymmetry. *Science* **306**, 1377–80 (2004).
166. Cenik, E. S. *et al.* Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. *Mol. Cell* **42**, 172–84 (2011).
167. Marques, J. T. *et al.* Loqs and R2D2 act sequentially in the siRNA pathway in *Drosophila*. *Nat. Struct. Mol. Biol.* **17**, 24–30 (2010).
168. Fukunaga, R. *et al.* Dicer partner proteins tune the length of mature miRNAs in flies and mammals. *Cell* **151**, 533–46 (2012).
169. Liu, X., Jiang, F., Kalidas, S., Smith, D. & Liu, Q. Dicer-2 and R2D2 coordinately bind siRNA to promote assembly of the siRISC complexes. *RNA* **12**, 1514–20 (2006).
170. Lee, J. Y. *et al.* Merlin, a tumor suppressor, interacts with transactivation-responsive RNA-binding protein and inhibits its oncogenic activity. *J. Biol. Chem.* **279**, 30265–73 (2004).
171. Yamashita, S. *et al.* Structures of the first and second double-stranded RNA-binding domains of human TAR RNA-binding protein. *Protein Sci.* **20**, 118–30 (2011).
172. Daniels, S. M. *et al.* Characterization of the TRBP domain required for dicer interaction and function in RNA interference. *BMC Mol. Biol.* **10**, 38 (2009).
173. Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557–80 (1983).
174. Fritsch; Sambrook & Maniatis. *Molecular Cloning: A Laboratory Manual SECOND EDITION*. (Cold Spring Harbor Laboratory, 1989). at <<http://www.amazon.com/Molecular-Cloning-Laboratory-Manual-EDITION/dp/B0019Y5D4O>>
175. Houben, K., Marion, D., Tarbouriech, N., Ruigrok, R. W. H. & Blanchard, L. Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. *J. Virol.* **81**, 6807–16 (2007).
176. Kapust, R. B. *et al.* Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **14**, 993–1000 (2001).
177. Vuister, G. W. & Bax, A. Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in ¹⁵N-enriched proteins. *J. Am. Chem. Soc.* **115**, 7772–7777 (1993).
178. Herrmann, T., Güntert, P. & Wüthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209–27 (2002).
179. Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* **104**, 4559–4570 (1982).

180. Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111–4 (1997).
181. Das, R. & Baker, D. Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **77**, 363–82 (2008).
182. Rule, G. & Hitchens, T. Fundamentals of protein NMR spectroscopy (Focus on structural Biology). *Dordr. Springer. xxvi*, 530 p (2006). at <<http://www.springer.com/series/5990>>
183. Reddy, T. & Rainey, J. K. Interpretation of biomolecular NMR spin relaxation parameters. *Biochem. Cell Biol.* **88**, 131–42 (2010).
184. Dosset, P., Hus, J. C., Blackledge, M. & Marion, D. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR* **16**, 23–8 (2000).
185. Palmer, A. G. NMR characterization of the dynamics of biomacromolecules. *Chem. Rev.* **104**, 3623–40 (2004).
186. d’Auvergne, E. J. & Gooley, P. R. The use of model selection in the model-free analysis of protein dynamics. *J. Biomol. NMR* **25**, 25–39 (2003).
187. Markley, J. L. *et al.* Recommendations for the presentation of NMR structures of proteins and nucleic acids--IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. *Eur. J. Biochem.* **256**, 1–15 (1998).
188. Lescop, E., Schanda, P. & Brutscher, B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.* **187**, 163–9 (2007).
189. Rückert, M. & Otting, G. Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *J. Am. Chem. Soc.* **122**, 7793–7797 (2000).
190. Herrmann, T. *Protein Structure Calculation and Automated NOE Restraints*. (John Wiley & Sons, Ltd, 2010).
191. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–23 (2009).
192. Güntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–78 (2004).
193. Rieping, W. *et al.* ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **23**, 381–2 (2007).
194. Zweckstetter, M. NMR: prediction of molecular alignment from structure using the PALES software. *Nat. Protoc.* **3**, 679–90 (2008).
195. Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time

- investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.* **127**, 8014–5 (2005).
196. Weigel, D. & Glazebrook, J. *Arabidopsis: A Laboratory Manual*. 354 (CSHL Press, 2002). at <<http://books.google.com.ar/books/about/Arabidopsis.html?id=IfZAMNPWV4C&pgis=1>>
197. Murashige, T. & Skoog, F. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiol. Plant.* **15**, 473–497 (1962).
198. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–54 (1976).
199. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–93 (1995).
200. Keller, R. *CARA (Computer Aided Resonance Assignment)*. (2004).
201. Johnson, B. A. & Blevins, R. A. NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–14 (1994).
202. Rohl, C. A. & Baker, D. De novo determination of protein backbone structure from residual dipolar couplings using Rosetta. *J. Am. Chem. Soc.* **124**, 2723–9 (2002).
203. Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 4685–90 (2008).
204. Simons, K. T. *et al.* Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins* **34**, 82–95 (1999).
205. Bonneau, R., Strauss, C. E. & Baker, D. Improving the performance of Rosetta using multiple sequence alignment information and global measures of hydrophobic core formation. *Proteins* **43**, 1–11 (2001).
206. Fogh, R. *et al.* The CCPN project: an interim report on a data model for the NMR community. *Nat. Struct. Biol.* **9**, 416–8 (2002).
207. Mandel, A. M., Akke, M. & Palmer, A. G. Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.* **246**, 144–63 (1995).
208. Delano, W. The Pymol Molecular Graphics System. (2002). at <<http://www.pymol.org/citing>>
209. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037–41 (2001).

-
210. Dosset, P., Hus, J. C., Marion, D. & Blackledge, M. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J. Biomol. NMR* **20**, 223–231 (2001).
211. Sivashanmugam, A. *et al.* Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Sci.* **18**, 936–48 (2009).
212. Masliah, G., Barraud, P. & Allain, F. H.-T. RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell. Mol. Life Sci.* **70**, 1875–95 (2013).
213. Stefl, R. *et al.* The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell* **143**, 225–37 (2010).
214. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9 (1998).
215. Rasia, R. M., Lescop, E., Palatnik, J. F., Boisbouvier, J. & Brutscher, B. Rapid measurement of residual dipolar couplings for fast fold elucidation of proteins. *J. Biomol. NMR* **51**, 369–78 (2011).
216. Hiraguri, A. *et al.* Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in *Arabidopsis thaliana*. *Plant Mol. Biol.* **57**, 173–88 (2005).
217. Nanduri, S., Rahman, F., Williams, B. R. & Qin, J. A dynamically tuned double-stranded RNA binding mechanism for the activation of antiviral kinase PKR. *EMBO J.* **19**, 5567–74 (2000).
218. Wostenberg, C., Quarles, K. A. & Showalter, S. A. Dynamic origins of differential RNA binding function in two dsRBDs from the miRNA “microprocessor” complex. *Biochemistry* **49**, 10728–36 (2010).
219. Wostenberg, C. *et al.* The role of human Dicer-dsRBD in processing small regulatory RNAs. *PLoS One* **7**, e51829 (2012).
220. Jacobsen, S. E., Running, M. P. & Meyerowitz, E. M. Disruption of an RNA helicase/RNase III gene in *Arabidopsis* causes unregulated cell division in floral meristems. *Development* **126**, 5231–43 (1999).
221. Bass, B. L., Hurst, S. R. & Singer, J. D. Binding properties of newly identified *Xenopus* proteins containing dsRNA-binding motifs. *Curr. Biol.* **4**, 301–14 (1994).
222. Jarymowycz, V. A. & Stone, M. J. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem. Rev.* **106**, 1624–71 (2006).
223. Boehr, D. D., Dyson, H. J. & Wright, P. E. An NMR perspective on enzyme dynamics. *Chem. Rev.* **106**, 3055–79 (2006).
224. Hartman, E. *et al.* Intrinsic dynamics of an extended hydrophobic core in the *S. cerevisiae*

-
- RNase III dsRBD contributes to recognition of specific RNA binding sites. *J. Mol. Biol.* **425**, 546–62 (2013).
225. Liu, C., Axtell, M. J. & Fedoroff, N. V. The helicase and RNaseIIIa domains of Arabidopsis Dicer-Like1 modulate catalytic parameters during microRNA biogenesis. *Plant Physiol.* **159**, 748–58 (2012).
226. Tagami, Y., Motose, H. & Watanabe, Y. A dominant mutation in DCL1 suppresses the hyl1 mutant phenotype by promoting the processing of miRNA. *RNA* **15**, 450–8 (2009).
227. Du, Z., Lee, J. K., Tjhen, R., Stroud, R. M. & James, T. L. Structural and biochemical insights into the dicing mechanism of mouse Dicer: a conserved lysine is critical for dsRNA cleavage. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 2391–6 (2008).
228. Weinberg, D. E., Nakanishi, K., Patel, D. J. & Bartel, D. P. The inside-out mechanism of Dicers from budding yeasts. *Cell* **146**, 262–76 (2011).