RIDAA
Repositorio Institucional
Digital de Acceso Abierto de la
Universidad Nacional de Quilmes

Universidad
Nacional
de Quilmes

Ormazábal, Agustín

Bases moleculares del mecanismo de regulación post-transcripcional asociado al sistema RsmZ/RsmE. Un estudio computacional

Esta obra está bajo una Licencia Creative Commons Argentina.
Atribución - No Comercial - Sin Obra Derivada 2.5
https://creativecommons.org/licenses/by-nc-nd/2.5/ar/

Documento descargado de RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes de la Universidad Nacional de Quilmes

Cita recomendada:
Ormazábal, A. (2024). Bases moleculares del mecanismo de regulación post-transcripcional asociado al sistema RsmZ/RsmE. Un estudio computacional. (Tesis de doctorado). Universidad Nacional de Quilmes, Bernal, Argentina. Disponible en RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes http://ridaa.unq.edu.ar/handle/20.500.11807/4279

Puede encontrar éste y otros documentos en: https://ridaa.unq.edu.ar

Universidad Nacional de Quilmes
Roque Sáenz Peña 352 // Bernal
Buenos Aires // Argentina
t.: (+41 11) 4365 7100
f.: (+54 11) 4365 7101
info@unq.edu.ar

Agustín Ormazábal, Repositorio Institucional Digital de Acceso Abierto,
Febrero de 2024, pp. 217
http://ridaa.unq.edu.ar,
Universidad Nacional de Quilmes, Secretaría de Posgrado,
Doctorado en Ciencia y Tecnología

Bases moleculares del mecanismo de regulación post-transcripcional asociado al sistema RsmZ/RsmE. Un estudio computacional

TESIS DOCTORAL

Agustín Ormazábal

agustin.ormazabal.5@gmail.com

Resumen

Las proteínas de la familia Csr se encuentran presentes en el 75% de las especies bacterianas, y se estima que el 15% de los genes codificados en *Escherichia coli* son regulados por éstas. Su rol biológico está asociado a su capacidad de unirse a las regiones no traducidas (UTR) 5' de ARN mensajeros (ARNm), modificando así su tasa de traducción. Entre los ARNm blanco de RsmE se destaca el gen *hcnA*, el cual codifica para una de las subunidades de la enzima responsable de sintetizar HCN en *Pseudomonas protegens*. Dicho metabolito actúa como un inhibidor natural frente a patógenos del suelo, por lo que las bacterias capaces de sintetizarlo proveen una protección natural a las plantas cuyas races habitan. Las homólogas de la familia Csr en el género *Pseudomonas* son nombradas con el prefijo "Rsm". Éstas impiden la accesibilidad de la unidad ribosomal 30s al gen objetivo, de manera que la presencia de la proteína implica una represión en la expresión del ARNm. La familia Csr/Rsm es a su vez regulada por moléculas de ARN no codificantes y de secuencia corta (ARNp), capaces de capturar a la proteína y liberar al gen sobre el cual ejercen una función regulatoria. Los ARNp en cuestión poseen la misma secuencia consenso de unión que los ARNm capturados por Csr/Rsm, actuando así como inhibidores competitivos. Entre estos ARNp, RsmZ es el único cuya estructura ha sido caracterizada experimentalmente. En dichos trabajos a su vez se determino que RsmE se une cooperativamente los distintos bolsillos de unión de RsmZ siguiendo un orden definido. Contrariamente, estudios experimentales previos demuestran que RsmE se une a *hcnA* de manera anticooperativa. El objetivo de esta tesis es explicar dichos aspectos del sistema. Para esto, se presenta un análisis basado en simulaciones de Dinámica Molecular (DM) que explora la diversidad conformacional tanto de RsmZ como de los complejos RsmE-RsmZ y RsmE-*hcnA*. Los resultados revelan un patrón definido de exposición secuencial de los diferentes motivos de unión de RsmZ causado por la captura de las sucesivas unidades de RsmE. En conjunto, estas simulaciones brindan una explicación simple y consistente para el orden de unión observado experimentalmente, y cuyas bases moleculares no habían sido

develadas hasta el momento. Además, se presentan modelos estructurales para los complejos RsmE-RsmZ parcialmente ocupados que no habían sido descritos de forma experimental al momento de realizarse esta tesis. Por otra parte, se implementaron simulaciones de *Umbrella Sampling* que describen el mecanismo de unión entre RsmE y el sitio de unión de RsmZ situado en una región simple cadena. El estudio estuvo enfocado en dilucidar cuáles son las interacciones más relevantes del proceso, estimando la energía involucrada en el evento de unión. Se observó que el motivo estudiado de RsmZ adquiere una estructura símil *stem-loop* durante el evento, y que para poder llegar a esa conformación es necesario que el segmento de simple cadena contenga al menos nueve nucleótidos. Asimismo, y mediante un Análisis de Componentes Principales (PCA) de ambas moléculas en su forma libre, se determinó que sus principales coordenadas colectivas tienden a generar las conformaciones propicias para la unión, sugiriendo fuertemente que la flexibilidad de la región monocatenaria de RsmZ afecta de manera crucial su capacidad para capturar a RsmE. Finalmente, se presentan resultados preliminares realizados para dilucidar el mecanismo de unión entre RsmE y el gen *hcnA*. Las simulaciones de *Umbrella Sampling* presentadas reproducen cabalmente los resultados experimentales y, por lo tanto, pueden ser luego aplicados a su racionalización. En conjunto, este trabajo presenta las primeras simulaciones reportadas tanto para el sistema de regulación post-transcripcional RsmE-RsmZ como para el complejo RsmE-*hcnA*, y se suma a los escasos antecedentes de técnicas de DM aplicadas al estudio de moléculas de ARNp.

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

Universidad Nacional de Quilmes

Bases moleculares del mecanismo de regulación post-transcripcional asociado al sistema RsmZ/RsmE

Un estudio computacional

Tesis doctoral
Doctorado en Ciencia y Tecnología
Universidad de Quilmes, diciembre 2023

Agustín Ormazábal
Dirección: Juliana Palma
Co-dirección: Gustavo Pierdominici-Sottile

Bases moleculares del mecanismo de regulación post-transcripcional asociado al sistema RsmZ/RsmE: un estudio computacional

Por Agustín Ormazábal

Dirección: Juliana Palma

Codirección: Gustavo Pierdominici-Sottile

UNIVERSIDAD NACIONAL DE QUILMES

DICIEMBRE DE 2023

BASES MOLECULARES DEL MECANISMO DE REGULACIÓN POST-TRANSCRIPCIONAL ASOCIADO AL SISTEMA RSMZ/RSME: UN ESTUDIO COMPUTACIONAL

TESIS DOCTORAL - DOCTORADO EN CIENCIA Y TECNOLOGÍA - UNIVERSIDAD NACIONAL DE QUILMES
CANDIDATO *Agustín Ormazábal*
DIRECCIÓN: *Juliana Palma*
CO-DIRECCIÓN: *Gustavo Pierdominici-Sottile*
Diciembre de 2023

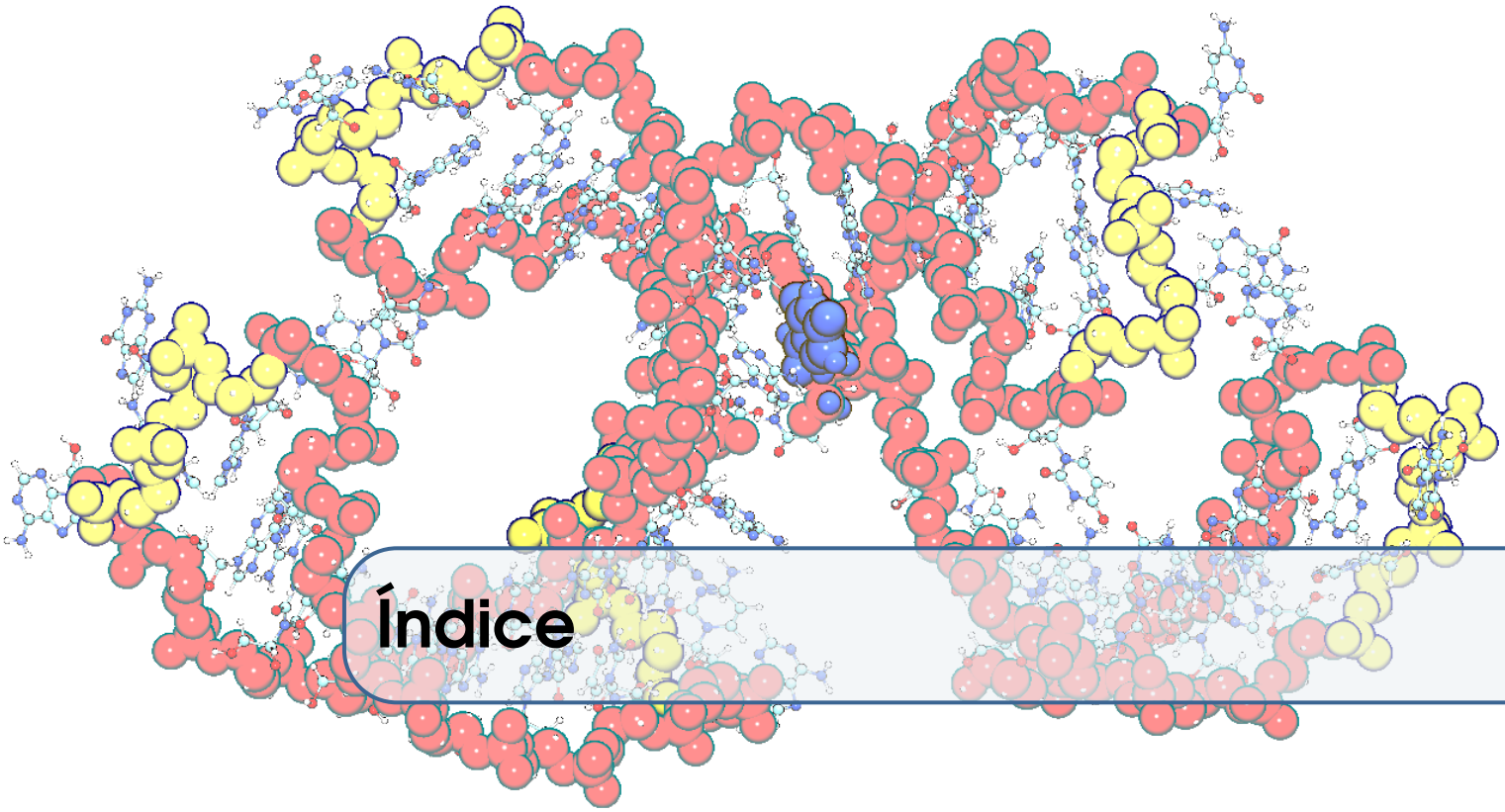
ARTE DE TAPA: PAULA ALBIN

Esta tesis se basa en una investigación realizada entre 2018 y 2023 en la Universidad Nacional de Quilmes. Dicha investigación fue financiada tanto por la Universidad como por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y la Agencia Nacional de Promoción Científica y Tecnológica (AGENCIA). El primer borrador de esta tesis comenzó a escribirse en 2020 durante la pandemia del SARS-CoV-2, en Bernal, al sur del Conurbano Bonaerense. Terminó de bosquejarse en Julio de 2023 en San Bernardo, Partido de la Costa, y volvió a Bernal para terminar de corregirse en Diciembre de 2023.

Sólo se escribe lo que no está,
lo que ya no queda,
lo que es necesario apuntar
porque se olvida.

- Gata Cattana -

Dedicada al Conurbano Bonaerense,
a la Educación Pública,
y a la memoria de quienes ahora escriben a través de mi.



Índice

I RsmE como reguladora de la expresión génica ... 1

I.1 La familia Csr/Rsm de proteínas reguladoras en bacterias 2

I.1.1 Proteínas homólogas de CsrA 3

I.1.2 Mecanismos de regulación mediados por CsrA 5

I.1.3 Características estructurales de CsrA 7

I.2 La familia de ARNp reguladores de Csr/Rsm 9

I.3 RsmE como regulador en la síntesis de HCN 16

II Estructura y dinámica de los ARN 21

II.1 Características estructurales de los ARN 22

II.1.1 Estructuras secundarias de los ARN 25

II.1.2 Estructuras terciarias de los ARN 31

II.1.3 Modularidad de ensambles y jerarquías estructurales de los ARN 33

II.2 Ejemplos de estudios computacionales sobre el ARN 41

II.2.1 Antecedentes sobre ARN pequeños (ARNp) 43

III Conceptos básicos de Dinámica Molecular (DM) 47
III.1 Fundamentos 48
III.2 El modelo y las trayectorias 49
III.3 Los campos de fuerza (CF) 54
III.3.1 Campos de fuerza para ARN 57
III.3.2 Las etapas de una simulación 60
III.4 *Umbrella sampling* 63
III.5 Análisis 65
III.5.1 Análisis de Componentes Principales 66
III.5.2 Clusterización basada en densidad 69

IV Sobre el mecanismo de captura RsmE de por RsmZ 76
IV.1 Construcción de los modelos computacionales 79
IV.2 Simulación del complejo en distintos estados de ocupación 80
IV.3 Comportamiento de RsmZ y de los complejos RsmZ-RsmE 81
IV.3.1 Clusterización y caracterización estructural 87
IV.3.2 Caracterización estructural de RsmZ y sus complejos con RsmE 91
IV.3.3 Accesibilidad de los motivos de unión en RsmZ 93
IV.4 Conclusiones del capítulo 102

V Sobre el reconocimiento y unión de RsmE por RsmZ 106
V.1 Construcción de los modelos y simulaciones de DM estándar 108
V.2 Simulaciones de *Umbrella Sampling* 111
V.3 Mecanismo de unión entre RsmE y RsmZ 113
V.4 Cambios conformacionales durante el proceso de unión 118
V.4.1 Deformaciones de RsmE 118
V.4.2 Deformaciones de RsmZ-$(RsmE)_2$ 120

V.5 Deformaciones de RsmE y RsmZ-$(RsmE)_2$ aisladas **122**
V.5.1 Análisis de Componentes Principales 125
V.5.2 Movimientos esenciales de RsmZ-$(RsmE)_2$ y RsmE 127
V.6 Conclusiones del capítulo **129**
V.7 Anexo **132**

VI Perspectiva: Mecanismo de unión de RsmE a *hcnA* 139
VI.1 Construcción de modelos **142**
VI.2 Simulaciones de *Umbrella Sampling* **142**
VI.2.1 Perfiles de Energía Libre 145
VI.3 Mecanismo de unión entre RsmE y sus ARN objetivo **147**
VI.4 Racionalización de resultados experimentales previos **148**
VI.5 Conclusiones del capítulo **149**
VI.6 Anexo **150**

VII Discusión y Conclusiones generales 158
VII.1 Agradecimientos **166**

Bibliografía 171

Preámbulo

Esta tesis está dividida en dos partes, cada una de las cuales cuenta con tres capítulos. El criterio por el cual establecí esta división es distinguir el marco teórico, plasmado en la primera parte, de la implementación metodológica propiamente dicha, que se presenta en la segunda parte. La segunda parte incluye además las conclusiones generales del trabajo, la bibliografía empleada, y una tabla que resume las siglas utilizadas a lo largo del texto.

El primer capítulo introduce al sistema estudiado en este trabajo, el cual se trata de un tándem de regulación post-transcripcional que opera sobre la expresión de ciertos genes bacterianos. Este sistema se conoce como RsmE-RsmZ, y está implicado en la síntesis de metabolitos secundarios que proveen una protección natural a las plantas cuyo entorno habitan las bacterias que lo contienen. Una de las principales motivaciones de este trabajo ha sido dilucidar las bases moleculares del funcionamiento del sistema RsmE-RsmZ, así como de los genes sobre los que ejerce una función de regulación, a sabiendas de que el conocimiento generado podría contribuir al desarrollo de factores de biocontrol naturales para un enfoque ecológico de agricultura sustentable. Aún en los pasajes de este trabajo en los que esa motivación no sea explícita, fue siempre ésta la que impulsó la elección del sistema y sus posteriores etapas de estudio.

El segundo capítulo se centra en las características dinámicas y estructurales de las moléculas de ARN. El motivo por el cual se dedica un capítulo entero a este aspecto radica en que, por un lado, este tipo de moléculas no son las más ampliamente estudiadas mediante técnicas de simulación computacional, pero además porque los fenómenos más detenidamente estudiados en este trabajo se relacionan, precisamente, con esas características propias de los ARN. El tercer capítulo discute los fundamentos teóricos de las simulaciones de Dinámica Molecular, siendo ésta la técnica sobre la que se basó este trabajo. Habiendo introducido ya las características del ARN, se discute cómo éstas afectan a la aplicación de técnicas de Dinámica Molecular, y a los métodos de análisis generalmente aplicados a proteínas. Parte de los temas discutidos en estos dos capítulos fueron recientemente incluidos en el libro “Bacterial Regulatory RNA: Methods and Protocols” de *Springer Nature* [1],

en el cual participó nuestro grupo.

Los tres capítulos de la segunda parte están divididos de manera que cada uno de ellos aborda objetivos generales y específicos distintos, así como hipótesis diferentes. Esta segunda parte comienza con el Capítulo IV, el cual explica cómo se aplicaron las técnicas de Dinámica Molecular, Análisis de Componentes Principales y métodos de clusterización para hallar un patrón estructural y dinámico que permitiera explicar el carácter secuencial y cooperativo a través del cual RsmE se une a RsmZ. Los resultados presentados en dicho capítulo fueron publicados en *The Journal of Physical Chemistry B* en 2021 [2].

El quinto capítulo aborda la simulación del evento de unión entre RsmE y RsmZ propiamente dicho. Asimismo, también se describieron las principales deformaciones sufridas por el sistema durante el proceso de unión, los residuos que juegan un papel crítico en ese mecanismo, y el perfil de energía asociado al evento. El análisis reveló que la longitud del segmento de ARN al que se une RsmE cumple un rol fundamental para posibilitar esa unión. Los resultados presentados en dicho capítulo fueron publicados en *Journal of Chemical Information and Modeling* en 2022 [3].

El sexto capítulo describe un nuevo proceso de unión, pero esta vez protagonizado por RsmE y un segmento de uno de los ARN mensajeros sobre los que ejerce una función regulatoria: el gen *hcnA*. Este gen codifica una de las unidades de la enzima responsable de sintetizar ácido cianhídrico, uno de los metabolitos secundarios de *Pseudomonas protegens* que actúan como inhibidores de crecimiento para patógenos de plantas. Dicho capítulo plantea la perspectiva de continuidad para esta línea de investigación, describiendo mediante técnicas de muestreo mejorado el carácter anti cooperativo en el tándem RsmE-*hcnA* y dejando sentadas las hipótesis que permitirían explicar dicho comportamiento.

Finalmente, las conclusiones generales de este trabajo aúnan los resultados obtenidos en cada una de sus etapas, aportando una mirada integral de los mismos en el contexto del conocimiento previo sobre el sistema, e intentando relacionar diversas observaciones experimentales que hasta el momento aparecían como inconexas. Los trabajos reportados en esta tesis implican los primeros antecedentes

de simulaciones abocadas al estudio del sistema de regulación post-transcripcional RsmE-RsmZ.

El documento es interactivo, de manera que tanto las referencias, como los títulos de las secciones y las menciones al material complementario, están hipervinculadas como enlaces. De esta manera, presionando sobre cada una de estas menciones en el cuerpo del texto se redirige al elemento referenciado, sea dentro del propio documento como fuera de él. En este sentido, cabe mencionar que la tesis cuenta con modelos estructurales disponibles en el repositorio https://sketchfab.com/aormazabal, y a lo largo del texto se proveen códigos QR que, al escanearse, permiten visualizar las estructuras en Realidad Aumentada. A su vez, se puede acceder a videos con animaciones de los procesos simulados a través del enlace: https://www.youtube.com/@agusin.ormazabal/playlists.

Cabe destacar que, en paralelo a esta investigación, realicé una estancia de investigación de seis meses en el Instituto Europeo de Bioinformática de Cambridge (EMBL-EBI) bajo la dirección de Alex Bateman. Durante este período, trabajé en la caracterización de proteínas repetitivas humanas con capacidad de unión a ARN. En particular, el trabajo estuvo centrado en distinguir las propiedades de dichas interacciones según si el tipo de repeticiones de las proteínas involucradas era de carácter secuencial o estructural. Si bien la temática es tangencial a la abordada durante mi doctorado, me permitió ampliar mi visión sobre las interacciones del tipo proteína-ARN, lo cual fue esencial para la escritura de este trabajo. Los resultados de dicha investigación fueron publicados en PLOS ONE en el año 2023 [4].

Respecto a la historia de esta tesis, la misma se remonta al año 2017. Todo empezó cuando nos encontramos con el que sería el sistema a estudiar, gracias a la gentileza de Claudio Valverde, docente e investigador de nuestra Universidad, quien nos puso al tanto de la reciente obtención de la estructura tridimensional del complejo RsmE-RsmZ. Durante el desarrollo de esta tesis se produjo un punto de inflexión en el campo de la biología y la biofísica computacional, como fue el advenimiento de las inteligencias artificiales para predecir estructuras de proteínas a partir de sus secuencias. Esas herramientas hoy empiezan a dar, con las dificultades del caso, sus primeros pasos en lo que respecta a la predicción estructural de ácidos

nucleicos, y en etapas posteriores (probablemente más fatigosas), seguramente hagan lo propio para los complejos formados por ambos tipos de moléculas. No obstante, esa perspectiva parecía muy lejana cuando este trabajo tuvo comienzo, por lo que contar con la estructura del complejo RsmE-RsmZ fue un insumo esencial para su estudio.

Para nuestro grupo avanzar con esta investigación implicó inaugurar una nueva línea de investigación sobre la que no teníamos experiencia. Implicó acostumbrarse a comprender que el ojo a través del cual hay que observar una simulación que involucra moléculas de ARN es completamente distinto al necesario para observar la dinámica de una proteína. Implicó adaptar las técnicas que hasta ese momento nos resultaban habituales a un nuevo contexto biológico. Y sobre todo, implicó mucha artesanía, paciencia y trabajo en equipo.

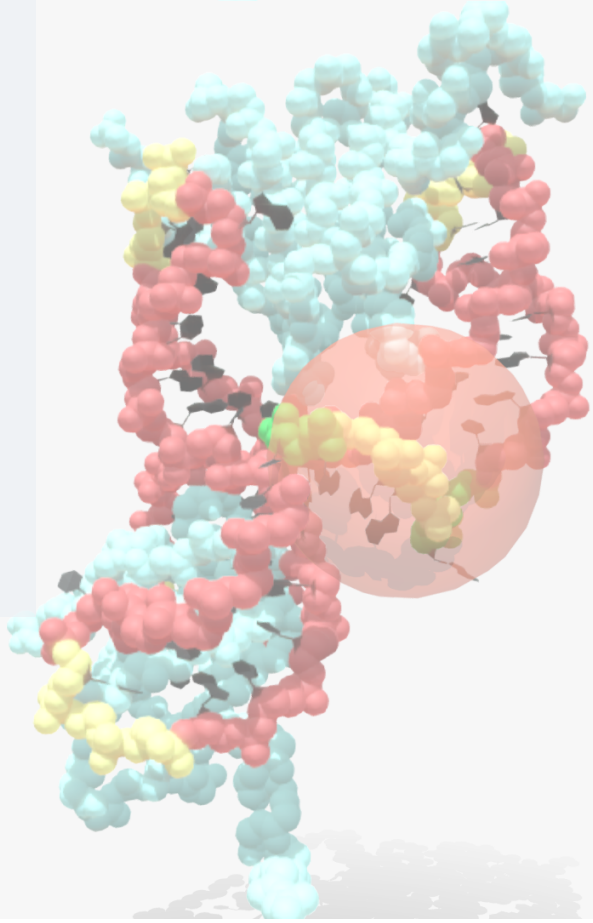
Podría decirse que la apuesta por investigar este tema fue, ante todo, bastante inconsciente. Nos ganó completamente el entusiasmo del momento, y creo hablar tanto por mi como por quienes dirigieron este trabajo al decir que, con los matices propios del paso del tiempo, esa motivación nos acompañó hasta el momento en el que cayó el último punto sobre esta tesis. Y hasta después, también.

Juliana, Gustavo y yo somos un equipo pequeño. Pero sin ese constante trabajo en equipo no existiría ninguno de los capítulos que suceden a esta pequeña introducción. Si bien me toca ser quien narra la historia de estos cinco años y medio de investigación, el rol que cada quién asumió para el desarrollo del trabajo es intercambiable. Sobre todo, a nivel personal. Esas tres dimensiones del equipo son la identidad del resultado final de ese trabajo que es esta tesis.

PARTE I

Marco teórico y estado del arte

I.1 La familia Csr/Rsm de proteínas reguladoras en bacterias 2
I.1.1 Proteínas homólogas de CsrA 3
I.1.2 Mecanismos de regulación mediados por CsrA 5
I.1.3 Características estructurales de CsrA . 7
I.2 La familia de ARNp reguladores de Csr/Rsm 9
I.3 RsmE como regulador en la síntesis de HCN 16



I. RsmE como reguladora de la expresión génica

Resumen de este capítulo Las proteínas de la familia Csr se encuentran presentes en el 75% de las especies bacterianas, y se estima que el 15% de los genes codificados en *Escherichia coli* son regulados por éstas. Su rol biológico está asociado a su capacidad de unirse a las regiones no traducidas (UTR) 5′ de ARN mensajeros (ARNm), modificando así su tasa de traducción. Las homólogas de CsrA de la familia Csr en el género *Pseudomonas* son nombradas con el prefijo "Rsm". Éstas impiden la accesibilidad de la unidad ribosomal 30s al gen objetivo, de manera que la presencia de la proteína implica una represión en la expresión del ARNm. La familia Csr/Rsm es a su vez regulada por moléculas de ARN no codificantes y de secuencia corta (ARNp), capaces de capturar a la proteína y liberar al gen sobre el cual ejercen una función regulatoria. Los ARNp en cuestión poseen la misma secuencia consenso de unión que los ARNm capturados por Csr/Rsm, actuando así como inhibidores competitivos. Entre estos ARNp, RsmZ es el único cuya estructura ha sido caracterizada experimentalmente. Esta información estructural constituyó un insumo fundamental para la realización de esta tesis. En este capítulo se introduce el sistema de regulación post-transcripcional RsmE-RsmZ, y se

exponen los principales interrogantes que se abordarán en la segunda parte de esta tesis. Por último, se describe cómo la síntesis de estos ARNp es también regulada en función de factores de contexto, y la influencia que éstos tienen sobre la traducción de metabolitos secundarios relevantes por su rol ambiental. Así, se introduce a la familia de genes *hcnABC*, de particular interés para este trabajo.

En las siguientes secciones se describen los aspectos esenciales del complejo de regulación post-transcripcional RsmE-RsmZ, así como de los ARNm regulados por el mismo. También se presentan las principales incógnitas que existían al momento de iniciarse este trabajo. Los capítulos que conforman la segunda parte de esta tesis buscan, precisamente, dar respuesta a algunas de esas preguntas.

I.1 La familia Csr/Rsm de proteínas reguladoras en bacterias

Durante décadas, la adaptación de las bacterias a los cambios en su entorno fue asociada a factores de regulación sobre la transcripción de determinados genes [5–7]. En particular, la mayoría de los estudios hasta principios de siglo coincidían en asociar esta regulación al control sobre el inicio de la transcripción y, en menor medida, con su terminación [8]. Sin embargo, los últimos años trajeron importantes avances en términos de la información fisiológica, bioquímica, genética y ómica disponible. En particular, se identificaron etapas de regulación que ocurren *a posteriori* de la transcripción [9], generando un gran auge de investigaciones referidas a los mecanismos de regulación post-transcripcional durante las últimas dos décadas. Si bien, en la mayoría de los casos, los mecanismos de regulación post-transcripcional están asociados al control fino de la expresión de ciertos genes, existen ejemplos en los que su rol es determinante para el ajuste de rutas metabólicas asociadas a factores contextuales [9]. Tal es el caso de la familia de proteínas Csr.

CsrA (por sus siglas en inglés *Carbon storage regulator A*) fue la primera proteína de la familia en ser identificada. Su descubrimiento se dio en el año 1993, como consecuencia de una serie de estudios basados en mutagénesis dirigida que tenían como objetivo la identificación de proteínas de *Escherichia coli* capaces de unirse de forma secuencia-específica a ácidos ribonucleicos (ARN) [10]. En ese

mismo trabajo se reportó que CsrA se trata de un polipéptido de 61 aminoácidos.

Al momento de descubrirse esta proteína, su secuencia no se correspondía con ninguna otra conocida. Por este motivo, no era posible inferir su función específica basándose en la secuencia. No obstante, en esos experimentos se observó que la deleción del gen codificante para CsrA (*csrA*) tenía consecuencias fenotípicas en las vías de flujo del Carbono de las bacterias [10]. El nombre de la proteína se debe precisamente a esa observación. En un estudio posterior realizado ese mismo año, se encontró que los mutantes deletéreos de *csrA* sufren alteraciones en su forma celular y su capacidad de adhesión a superficies, entre otros efectos pleiotrópicos [11]. Hasta la fecha se ha reportado que CsrA participa de mecanismos involucrados con un espectro amplio de funciones biológicas que incluyen la motilidad celular, la virulencia, la síntesis de glucógeno, el *quorum sensing* y las respuestas a estrés, entre otras [8, 12–17].

El descubrimiento de CsrA abrió la puerta a nuevos estudios abocados a su caracterización biológica. En el año 1996 se abordó por primera vez la distribución filogenética de la proteína a partir de la búsqueda de homologías secuenciales por *Southern Blot*, partiendo de amplificaciones por PCR (por *Polymerase Chain Reaction*) [18]. Aunque esta metodología alcanzó un número limitado de genomas bacterianos, arrojaron un primer indicio sobre la amplia distribución de esta proteína entre eubacterias.

I.1.1 Proteínas homólogas de CsrA

La primera homóloga de CsrA fue descubierta en la γ-proteobacteria *Pectobacterium carotovorum*, y se denominó RsmA por sus siglas en inglés *Repressor of Secondary Metabolites A* [19]. Su nombre se debe a su efecto represor en hospedadores patogénicos través de la regulación de enzimas líticas y del mecanismo de *quorum sensing*. Si bien la subfamilia Rsm presenta a grandes rasgos diferencias importantes desde el punto de vista de sus secuencias, posee una alta similitud estructural con el resto de la familia Csr [9, 20]. Por simplicidad, las proteínas de la familia Csr y la subfamilia Rsm serán mencionadas en adelante como la familia “Csr/Rsm”, dado que a fines de esta tesis representan un conjunto

de proteínas estructuralmente equivalentes.

Está descrito que la distribución filogenética de estas proteínas no es homogénea, sino que se encuentran ausentes en especies que poseen ciclos de vida intracelulares, como *Rickettsia, Chlamydia* y la γ-proteobacteria *Francisella* [9]. En ese mismo trabajo se muestra que incluso dentro de la rama proteobacteria algunas clases carecen de homólogos de Csr/Rsm, mientras en bacterias tales como las β-proteobacterias su presencia se limita a las especies *Bordetella petrii* y *Burkholderia pseudomallei* (cepa TSV202). Este mismo estudio destaca que *Bordetella petrii* no cumple un rol patogénico, sino que ejerce una función ambiental. Este es un aspecto compartido con *Pseudomonas aeruginosa*, siendo esta última una de las especies que expresa esta familia de proteínas.

Las homólogas de CsrA en el género *Pseudomonas* son de particular interés para este trabajo. Entre dichas homólogas, RsmA es la más ampliamente distribuida [9]. Casi todos los genomas del género *Pseudomonas* presentan un alelo para el gen *rsmA*. En segundo lugar, se ubica la subfamilia RsmE, presente en las especies *chlororaphis, corrugata, fragi, koreensis, lutea, mandelii, putida, syringae* y, de particular relevancia para este trabajo, *protegens*. RsmC y RsmI son las otras dos homólogas de CsrA muy extendidas dentro del género *Pseudomonas*, mientras el resto de las subfamilias (RsmN, RsmM, RsmD, RsmL y RsmH) tienen una distribución mucho más restringida [9]. La tabla I.1.1 resume los nombres de CsrA y sus homólogas en diferentes organismos del género *Pseudomonas*.

Al día de hoy, se estima que CsrA y sus homólogas interactúan con el 15% de los genes de *Escherichia coli* [21–23] y que las mismas están presentes en el 75% de las especies bacterianas [13]. Esto equivale a alrededor de 16.000 secuencias polipeptídicas asociadas a CsrA, las cuales se encuentran almacenadas en las bases de datos de Pfam e InterPro, abarcando cerca de 2900 especies [9]. Estas observaciones muestran la relevancia de las funciones ejercidas por la familia de proteínas Csr. Puede verse además que, entre estas funciones, su participación en mecanismos de regulación supera con creces al alcance inicialmente asignado a la proteína.

Tabla I.1: Resumen de los nombres adoptados por las proteínas homólogas de CsrA según la especie. Información recopilada de la Referencia 9.

Especie	Homóloga de CsrA
Acinetobacter baumannii	CsrA
Azotobacter vinelandii	RsmA
Erwinia amylovora	RsmA
Escherichia coli	CsrA
Halomonas anticariensis	CsrA
Legionella pneumophila	CsrA
Pectobacterium carotovorum	RsmA
Pseudomonas aeruginosa	RsmA, RsmN (RsmF)
Pseudomonas brassicacearum	RsmA, RsmE
Pseudomonas chlororaphis	RsmA, RsmE
Pseudomonas donghuensis	RsmA, RsmE, Rsm3
Pseudomonas entomophila	RsmA1, RsmA2, RsmA3
Pseudomonas fluorescens	RsmA, RsmE
Pseudomonas protegens	**RsmA, RsmE**
Pseudomonas putida	RsmA, RsmE, RsmI
Pseudomonas syringae	CsrA1-5, (RsmI, RsmA, RsmE)
Salmonella enterica serovar Typhimurium	CsrA
Serratia mascescems	CsrA
Vibrio cholerae	CsrA
Vibrio fischeri	CsrA
Vibrio tasmaniensis	CsrA

I.1.2 Mecanismos de regulación mediados por CsrA

La enzima ADP-glucosa pirofosforilasa, responsable de la síntesis de glucógeno, está codificada por el gen *glgC*. Este gen fue uno de los primeros en ser identificado como blanco de CsrA [24]. Para esto se usó un constructo entre *glgC* y *lacZ*, y se observó que su regulación no dependía de que existiera un promotor nativo para *glgC*. Esto sugirió que la regulación debía ocurrir de manera post-transcripcional. Años después esta suposición fue demostrada, encontrándose que la unión de CsrA a la secuencia *leader* de *glgC* bloquea el acceso del ribosoma al Sitio de Unión al Ribosoma (RBS, por sus siglas en inglés), inhibiendo así su traducción [25].

Posteriormente se encontró el mismo efecto en otros genes de *E. coli* [23, 26, 27]. También se han descrito contextos dentro de los cuales CsrA actúa estabilizando la formación de estructuras secundarias en el ARN blanco que luego impiden su traducción [28]. Un caso interesante es el del gen *pgaA*, necesario para la síntesis de la adhesina y la posterior formación del *biofilm*. En este caso, se observó que CsrA altera la estructura secundaria del ARNm, permitiendo el acceso del complejo Rho que eventualmente producirá el fin de la transcripción [29]. Este fue el primer caso descrito en el que CsrA regula la transcripción de manera directa.

También se han encontrado casos en los cuales CsrA actúa como activador positivo de la traducción. Se ha demostrado que el gen *moaA*, necesario para la biosíntesis del cofactor de Molibdeno (MOCO), es regulado positivamente cuando dos sitios de unión a CsrA ubicados en su Región no Codificante 5' (UTR 5', por sus siglas en inglés) son ocupados por dicha proteína. Esta interacción hace que el gen adopte una estructura óptima para la accesibilidad al ribosoma, activando así su traducción [30]. Asimismo, se ha descripto que CsrA regula positivamente la traducción de *flhDC*, responsable de controlar la motilidad de *Escherichia coli*, al evitar la degradación del ARN en cuestión [31, 32].

Los mecanismos mediante los que Csr/Rsm regula la expresión génica se resumen en el siguiente listado [9]:

- Represión de la traducción por unión competitiva al RBS.
- Activación de la traducción por rearreglo estructural de la UTR 5' y exposición del RBS.
- Regulación traduccional por replegamiento de la UTR 5', favoreciendo el ingreso de otras moléculas reguladoras de la traducción.
- Modulación de la concentración de ARNm por control del acceso de ribonucleasas.
- Modulación sobre la terminación de la transcripción.

Una característica que tienen en común todas las proteínas de la familia Csr/Rsm es la secuencia de consenso que reconocen en sus ARNm objetivo, así como también las estructuras que tienen los sitios de interacción con estas secuencias. La siguiente sección aborda esta temática.

I.1.3 Características estructurales de CsrA

Las primeras características estructurales descritas de CsrA y sus homólogas se basaron en predicciones de su estructura secundaria [33]. Las estructuras así predichas indicaban la presencia de cinco láminas-β consecutivas y antiparalelas, seguidas de una hélice-α corta y una región C-terminal desestructurada. Recién en el año 2005 fue resuelta la estructura de CsrA [34], y al año siguiente, se obtuvo la de RsmA de *Yersinia enterocolitica*. En ambos casos se pudo observar que CsrA es un homodímero cuyos monómeros están entrelazados a través de las láminas β que forman un centro hidrofóbico, con dos hélices α extendiéndose hacia las regiones C-terminal. Dicha forma de ensamblado es transversal a toda la familia Csr/Rsm [9]. La Figura I.1 muestra una representación pictórica de esta estructura.

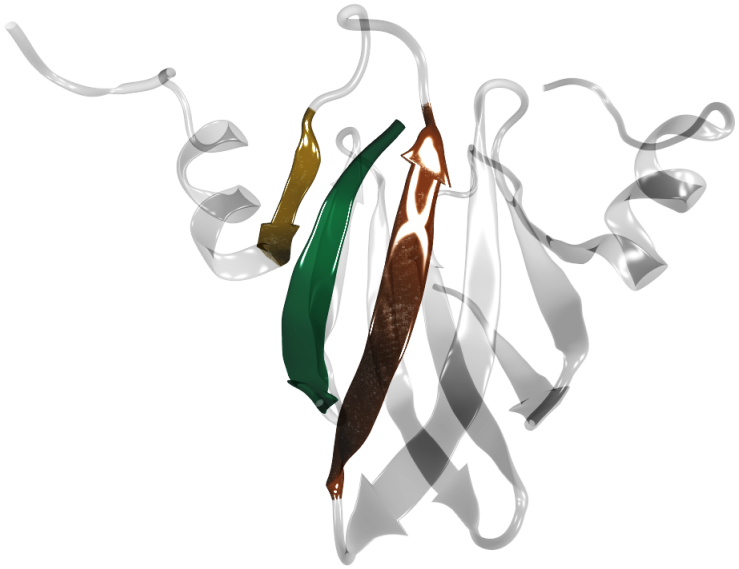


Figura I.1: Representación de la estructura de CsrA. Se resalta en verde a la lámina $\beta 1$ de una de las cadenas, así como en rojo y amarillo a las láminas $\beta 4$ y $\beta 5$ de la cadena complementaria, respectivamente. Dichas regiones de la proteína son las que establecen las principales interacciones con sus ARN objetivo.

En estudios posteriores que utilizan la técnica de mutagénesis por barrido de Alanina (en inglés *alanine-scanning mutagenesis*), se determinó que cada dímero puede interactuar con hasta dos moléculas de ARN a través de la superficie for-

mada por la lámina $\beta 1$ de una de las cadenas, y las láminas $\beta 4$ y $\beta 5$ de la cadena complementaria [35]. Esto se debe a la presencia de residuos cargados positivamente en dichas láminas. Entre estos residuos, se ha determinado que la Arginina altamente conservada en la interfaz entre $\beta 5$ y la hélice H1 es fundamental para el reconocimiento del primer residuo G en motivos GGA del ARN objetivo [26, 36]. Estos fueron los primeros indicios de una secuencia consenso de unión para los miembros de la familia Csr/Rsm.

Dado que CsrA forma un dímero simétrico, en su momento ya se había especulado con que sería capaz de unir simultáneamente dos sitios de una misma molécula de ARN, o bien dos moléculas distintas al mismo tiempo. En 2007 fue dilucidada la estructura de la proteína homóloga de CsrA en la cepa CHA0 de la rizobacteria *Pseudomonas protegens*, unida a dos fragmentos de la secuencia RBS del ARNm *hcnA* [26] (Ver Sección I.3). Esta estructura ratificó que el esqueleto de la proteína reconoce al ARN a través de una secuencia de tres nucléotidos (GGA) dentro de un *loop* con la secuencia consenso A(X)GGAX. En esta secuencia, la letra "X" indica que cualquier nucleótido puede ocupar ese lugar, mientras que el paréntesis implica que ese nucleótido puede estar presente o no.

Estudios posteriores de Evolución Sistemática de Ligandos por Enriquecimiento Exponencial (SELEX, por sus siglas en inglés) permitieron identificar moléculas de ARN candidatas a poseer zonas de unión a CsrA, el 100% de las cuales tenían al menos una secuencia compatible con la secuencia de consenso reportada en la Referencia 26 [37]. Se encontró que el 92% de dichos motivos estaban en regiones de simple-cadena, generalmente ubicados sobre los extremos de *stem-loops* (Ver Sección II.1.1 para más información sobre este tipo de estructuras del ARN). De hecho, mutaciones sitio dirigidas que modifican la estabilidad de los *stem-loops* mostraron una disminución en la afinidad de CsrA por los ARN identificados.

Las concentraciones citosólicas de CsrA aumentan entre 10 mil y 30 mil dímeros por célula durante el estadío de crecimiento en *batch* [38]. Esto da cuenta de la necesidad de un mecanismo de control que permita mantener o modificar los niveles citosólicos de CsrA en función del contexto metabólico del organismo. La estructura del dímero puede servir como andamio para interacciones proteína-ARN

o proteína-proteína, ambas impactando en la expresión génica. Por ejemplo, en ciertos linajes, CsrA coexiste con FliW, un regulador de la expresión génica flagelar y primera proteína en describirse como antagonista de CsrA [39].

En la bacteria Gram-positiva *Bacillus subtilis*, FliW está implicada en el mecanismo de regulación para la síntesis de flagelina, de manera que controla tanto sus niveles en citoplasma como su organización dentro del mismo [40, 41]. Esto se debe a que FliW puede unirse a CsrA de forma no competitiva, inhibiendo su acción represora sobre la síntesis de flagelina [42]. En 2016 fue resuelta la estructura cristalográfica de CsrA unida a FliW [43]. Este trabajo demostró que una región de FliW rica en aminoácidos de carga negativa se ubica próxima a la región de CsrA que interactúa con sus ARNm objetivo. A raíz de esta observación se especula con que la unión de FliW a CsrA desfavorece su interacción con el ARNm objetivo, debido a la repulsión provocada por la región negativamente cargada de FliW hacia el esqueleto del ARN.

En el estudio de la Referencia 43 se demuestra que la extensión de la secuencia C-terminal de CsrA, involucrada en la interacción con FliW, está sujeta a co-evolución con dicha proteína. De esta manera, esta región de CsrA sólo está presente en las bacterias que expresan FliW. Esto evidencia que otras moléculas fueron seleccionadas por el curso evolutivo para ejercer el rol de FliW. Estas moléculas fueron, precisamente, de ARN.

I.2 La familia de ARNp reguladores de Csr/Rsm

El reemplazo de una proteína por un ARN para asumir el rol de controlar la acción de Csr/Rsm implica un ahorro en términos tanto de tiempo como de la energía requerida para el proceso de síntesis de la molécula reguladora. Además, supone un aumento de la eficiencia de la unión a su antagonista, porque admite la posibilidad de sumar más sitios de interacción a Csr/Rsm. Estos sitios mimetizan a aquellos presentes en los ARNm objetivo dado que poseen las mismas secuencias consenso de unión a la proteína, lo que les permite actuar con un mecanismo de inhibición competitiva [27, 30, 44–46]. Una única molécula de ARN puede poseer varias secuencias consenso, capturando así un número mayor de unidades proteicas que

su versión ancestral FliW.

Los ARN que ejercen funciones de regulación sobre las proteínas Csr/Rsm son de secuencia corta o "pequeños" (ARNp) y de tipo no codificante (es decir, que ejercen su función *per sé* sin ser traducidos) [15]. CsrB fue el primer ARNp caracterizado como regulador de CsrA [47, 48]. Se ha demostrado que mutaciones realizadas sobre *csrB* afectan la expresión de genes dependientes de CsrA, y que su deleción tiene efectos pleiotrópicos en la fisiología de *Escherichia coli* [8].

El descubrimiento de CsrB permitió identificar las características que debe tener un ARNp para reconocer a las proteínas Csr/Rsm: Utilizando un algoritmo estocástico de plegamiento de ARN, se estableció que la estructura de CsrB contenía 15 *stem-loops*, la mayoría de los cuales tenía un motivo GGA ubicado en las regiones simple-cadena en el extremo de cada *loop* [49]. Estos motivos se encontraban flanqueados por nucleótidos conservados, siguiendo la secuencia consenso presentada en la sección anterior.

Los ARNp CsrC y McaS aportan regulaciones adicionales sobre CsrA cuando los niveles intracelulares de la proteína aumentan, de manera que los tres actúan de manera concertada para ejercer su mecanismo regulatorio en el organismo que los expresa [50, 51]. Si bien esto podría interpretarse como una redundancia, estudios realizados en la última década demuestran que condiciones distintas de crecimiento *in vivo* y factores regulatorios como la proteína Hfq inciden de forma diferente sobre la expresión de CsrB, CsrC y McaS. Esto sugiere que cada uno de estos ARNp tiene además distintos roles regulatorios [52–54].

Los mecanismos de regulación de las proteínas Csr/Rsm basados en ARNp están altamente conservados en γ-protobacterias [12]. En estos organismos, los nombres de estos ARNp llevan también el prefijo "Rsm", y su interacción con RsmE conduce al "secuestro" de la proteína. Esto posibilita la traducción del gen en cuestión por permitir el ingreso de la subunidad ribosomal 30s [31, 32, 55–62]. La Figura I.2 presenta un esquema de este mecanismo. Si bien las concentraciones citosólicas de estos ARNp suelen ser relativamente bajas, las mismas se compensan por su capacidad de capturar varias proteínas de manera simultánea. La Tabla I.2 resume cuáles son los ARNp reguladores de las proteínas Csr/Rsm según la especie en la

que se encuentran.

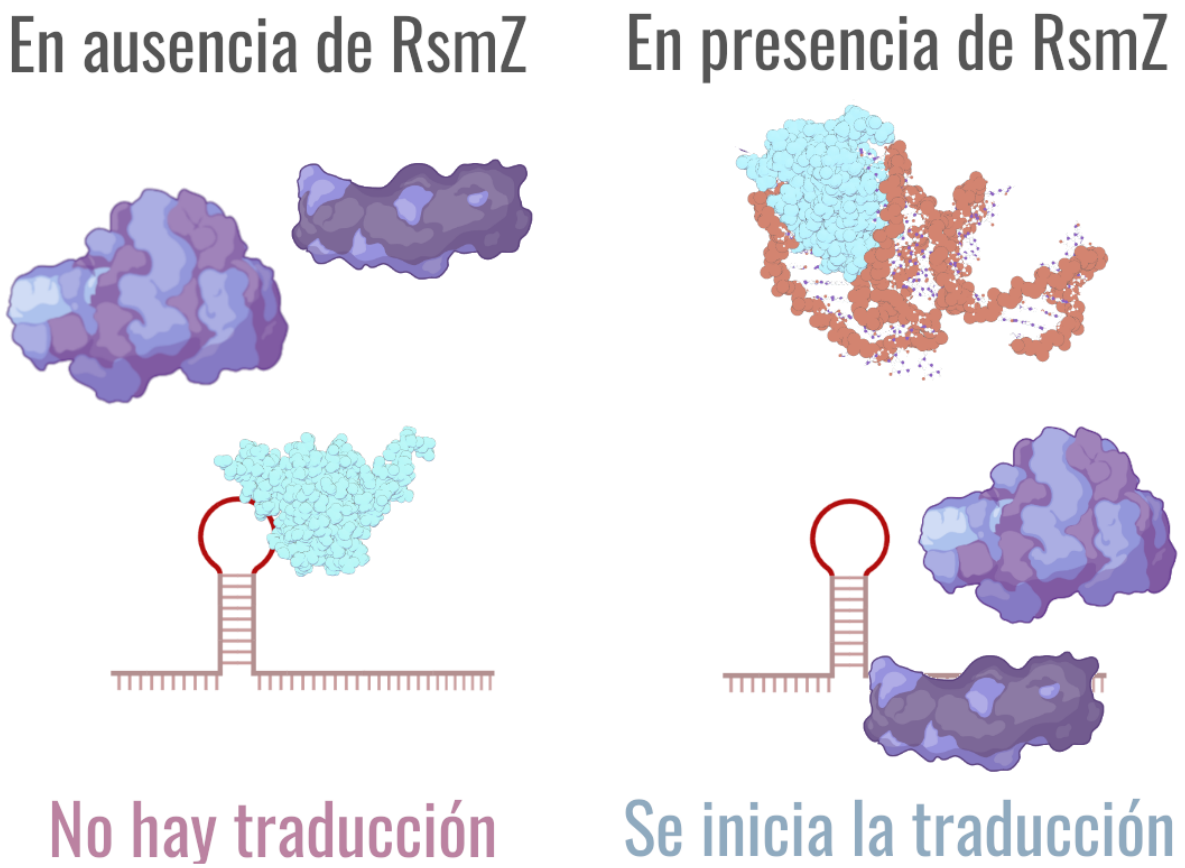


Figura I.2: Representación pictórica del comportamiento de RsmZ como activador de la traducción. En el panel izquierdo se muestra cómo en ausencia de RsmZ, RsmE (celeste) es capaz de unirse al ARNm objetivo (bordó) e impedir el ingreso de la unidad ribosomal 30s (violeta), inhibiendo la traducción del gen. En el panel derecho se muestra el caso contrario, donde la presencia de RsmZ (rojo) implica el secuestro de RsmE y la posibilidad de acceso para la subunidad ribosomal 30s al ARN objetivo, activando su traducción.

En las γ-proteobacterias, RsmX, RsmY y RsmZ son los los ARNp que están mejor caracterizados. Estos actúan de manera concertada para la regulación de las proteínas Rsm [63]. En la cepa CHA0 de *Pseudomonas protegens* se determinó que el complejo RsmE-RsmZ suprime la expresión de genes que codifican para factores de biocontrol [64–71]. De la misma manera, está descripto que este mecanismo

Tabla I.2: Resumen de los ARNp reguladores de las proteínas Csr/Rsm según la especie en la que se encuentran. Información recopilada de la Referencia 9.

Especie	ARNp regulador de CsrA
Acinetobacter baumannii	RsmX, RsmY, RsmZ
Azotobacter vinelandii	RsmZ1-7, RsmY1-2
Erwinia amylovora	RsmB
Escherichia coli	CsrB, CsrC
Halomonas anticariensis	Desconocido
Legionella pneumophila	RsmY, RsmZ
Pectobacterium carotovorum	RsmB
Pseudomonas aeruginosa	RsmY, RsmZ, RsmW,
Pseudomonas brassicacearum	RsmX, RsmY, RsmZ
Pseudomonas chlororaphis	RsmX, RsmY, RsmZ
Pseudomonas donghuensis	RsmY, RsmZ
Pseudomonas entomophila	RsmY, RsmZ
Pseudomonas fluorescens	RsmY, RsmZ
Pseudomonas protegens	**RsmX, RsmY, RsmZ**
Pseudomonas putida	RsmY, RsmZ, RsmX?
Pseudomonas syringae	CsrA1-5, (RsmX1-5, RsmY,RsmZ)
Salmonella enterica serovar Typhimurium	CsrB, CsrC
Serratia mascescems	CsrB, CsrC
Vibrio cholerae	CsrB, CsrC, CsrD
Vibrio fischeri	CsrB, CsrC, CsrD
Vibrio tasmaniensis	CsrB1-4

regulatorio juega un rol esencial en los factores de virulencia de numerosas plantas y mamíferos [8]. RsmZ es el único ARNp de *Pseudomonas protegens* cuya estructura se conoce. En primer lugar se resolvió la estructura de los complejos formados por RsmE y cinco porciones de RsmZ de la cepa CHA0 de *Pseudomonas protegens* [72], cada una de las cuales contenía un motivo de unión diferente. El estudio demostró que tanto la secuencia como las variaciones en el contexto estructural de los motivos de unión modulan la afinidad de dichos fragmentos por RsmE, pudiendo cambiar su constante de asociación/disociación en hasta cinco órdenes de magnitud [72].

Posteriormente, se empleó un enfoque combinado basado en espectroscopía de

Resonancia Paramagnética de Electrones (EPR) y Resonancia Magnética Nuclear (RMN) para caracterizar el mecanismo de unión de varios dímeros de RsmE a RsmZ. Esto permitió dilucidar la estructura de los primeros 72 nucleótidos de RsmZ unidos a tres unidades de RsmE [73]. A su vez, se describió que RsmZ forma un empalme de cuatro vías (ver Sección II.1.1) capaz de adoptar dos conformaciones, denominadas L y R. La Figura I.3 muestra la estructura secundaria de RsmZ y compara las estructuras terciarias de sus dos conformaciones. Las estructuras secundarias fueron determinadas haciendo uso del programa DSSR[74], mientras la aplicación VARNA fue empleada para graficarlas [75].

En cuanto al mecanismo de formación del complejo, el estudio pudo dilucidar que los motivos de unión de RsmZ se ocupan siguiendo un órden específico. El primer dímero de RsmE se adhiere a los motivos ubicados en los *stem-loops* (SL) 2 y 3. El segundo es capturado por los motivos ubicados en SL1 y SL4, mientras que el tercero se une al motivo situado en la región de cadena simple que conecta SL2 con SL3. Por lo tanto, en el complejo cuya estructura pudo ser resuelta, el tercer dímero está unido por un solo lado. El motivo de RsmZ que captura el lado restante se encuentra más allá del nucleótido 72, y su estructura no fue resuelta en dicho trabajo. La Figura I.4 ilustra el orden que sigue RsmE al unirse a la conformación R de RsmZ. El mismo orden es seguido por la conformación L.

Dado que el orden seguido por RsmE para ocupar los sitios de unión de RsmZ es siempre el mismo, se postuló un mecanismo de acción cooperativo. De esta manera, la unión de la primera unidad aumenta la afinidad del segundo bolsillo de unión, y así sucesivamente hasta que se ocupan todos los sitios del ARNp. Ese aspecto en particular es retomado en el Capítulo IV. Otro resultado interesante de ese trabajo fue la evaluación de las constantes de disociación desde los sitios de unión de RsmZ, las que resultaron diferentes entre sí y distintas de lo que cabría esperarse de acuerdo a las constantes de disociación de los fragmentos de RsmZ que fueron reportadas en la referencia 72.

Recientemente se descubrieron dos nuevos ARNp con capacidad de unión a CsrA en *Pseudomonas aeruginosa*: RsmV [76] y RsmW [77]. No obstante, se desconocen aún los detalles sobre su estructura tridimensional y las bases moleculares de su

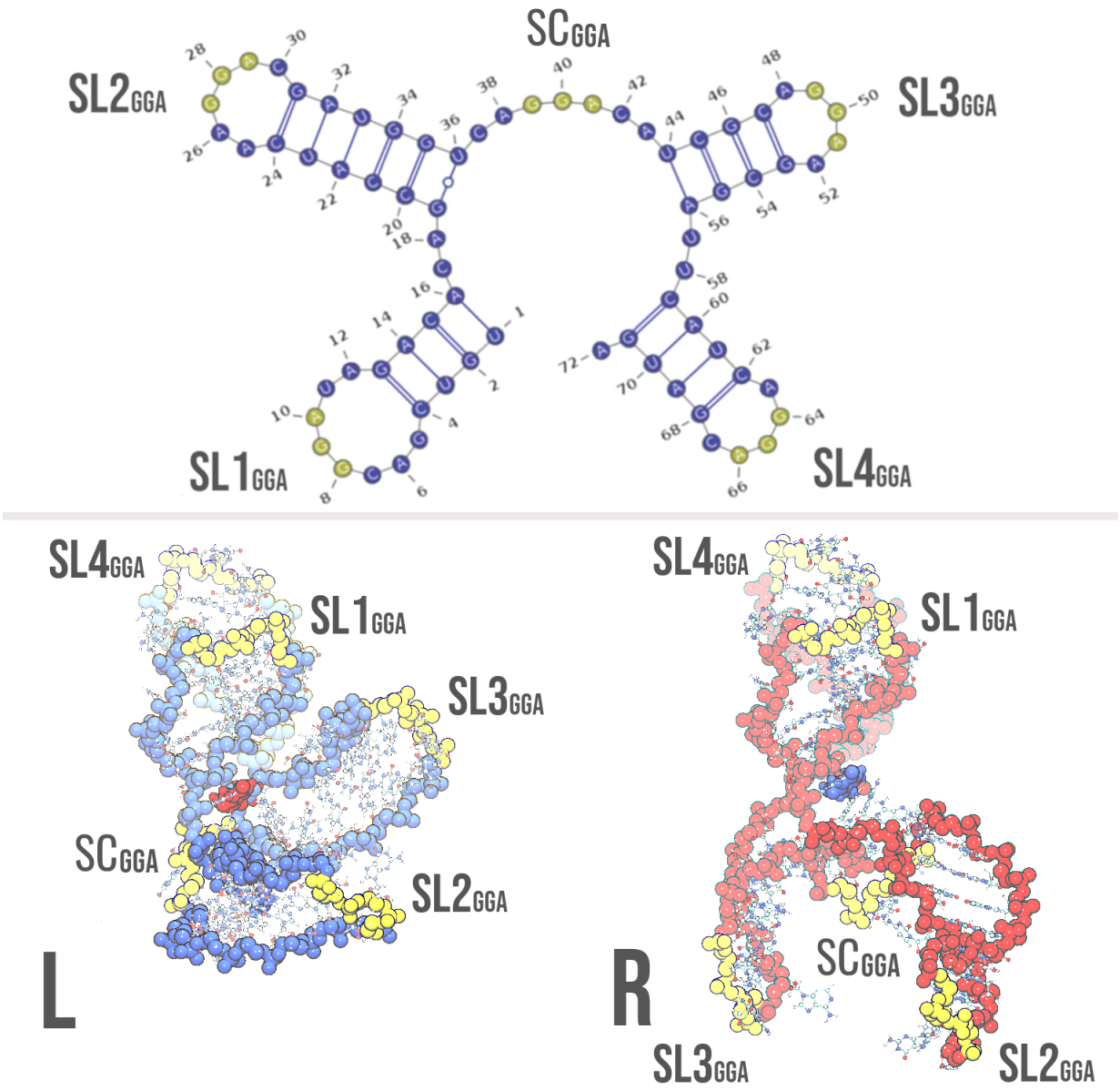


Figura I.3: **Panel superior**: estructura secundaria de RsmZ, común a ambos confórmeros. **Panel inferior**: Estructura terciaria de los confórmeros L y R. El esqueleto de la conformación L está mostrado en azul, excepto por el residuo U1 que aparece en rojo. El confórmero de la conformación R está coloreado en rojo, excepto el residuo U1, que aparece en azul. En ambos casos, los motivos GGA están destacados en amarillo. La imagen está adaptada de la Referencia 2.

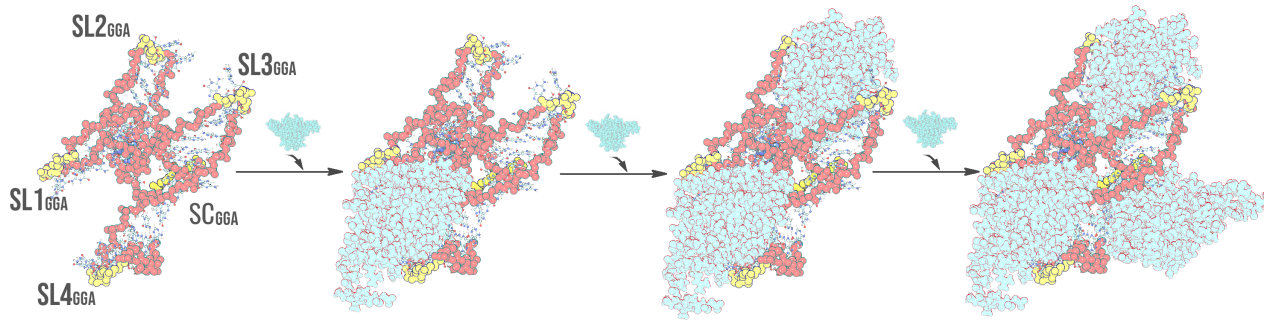


Figura I.4: Representación pictórica del orden seguido por RsmE (celeste) al unirse secuencial y cooperativamente a los motivos A(X)GGAX (amarillo) de RsmZ (cuyo esqueleto se marca en rojo). Esta representación corresponde a la conformación R del ARNp. Los nombres de cada motivo de unión se muestran en la estructura de RsmZ sin la proteína. La imagen está adaptada de la Referencia 2.

interacción con RsmE. El último hito ligado al sistema RsmE-RsmZ ocurrió en el año 2023, cuando dos nuevas estructuras de este complejo fueron reportadas [78]. Las mismas fueron obtenidas por criomicroscopía electrónica (Cryo-EM), y constituyen el primer antecedente de una estructura que contiene la secuencia completa de RsmZ, unida a dos y tres unidades de RsmE. Este trabajo coincide con el de la Referencia 73 respecto a cuál es el primer sitio de RsmZ ocupado por RsmE, pero difiere en el orden de ocupación de los sitios siguientes. Este aspecto en particular es retomado en los Capítulos IV y VI.

Los estudios computacionales del sistema RsmE-RsmZ son más bien escasos en comparación con los experimentales. Cálculos de Dinámica Molecular (DM) fueron implementados para evaluar la contribución de los diferentes nucleótidos de cada motivo GGA a la Energía Libre de unión del complejo [79]. Asimismo, estrategias de *docking* combinadas con simulaciones de DM fueron empleadas para estudiar los modos de unión de inhibidores de RsmE [80]. Sin embargo, hasta el inicio de este trabajo, no existían antecedentes de simulaciones sobre las estructuras completas del complejo RsmE-RsmZ.

Las concentraciones citoplasmáticas de los ARNp reguladores de las proteínas Csr/Rsm deben ser también controladas. En γ-proteobacterias, la expresión de estos

ARNp depende de la respuesta a estímulos ambientales censados por proteínas de membrana [8]. Así, el efecto inmediato es la fosforilación de un factor transcripcional asociado, que regula la expresión del gen o genes que codifican el ARNp en cuestión [9]. Este sistema de dos componentes se denomina GacS/GacA en bacterias del género *Pseudomonas* [8]. Gac (por sus siglas en inglés *Global Antibiotic and Cyanide control*) fue descubierta de manera simultánea con la ortóloga RsmA de CsrA [81]. Mientras la quinasa sensora GacS responde a señales químicas del entorno relacionadas con la densidad celular (al momento, químicamente desconocidas), GacA actúa como el factor intermediario para activar la transcripción de RsmX, RsmY y RsmZ [9]. La expresión de estos ARNp regula positivamente la síntesis de metabolitos secundarios como el ácido cianhídrico (HCN). La siguiente sección aborda el efecto de este sistema de regulación sobre el *cluster hcnABC*, responsable de la síntesis de dicho ácido.

I.3 RsmE como regulador en la síntesis de HCN

En bacterias del género *Pseudomonas*, GacA ejerce una función de regulación positiva sobre la síntesis de HCN, así como de otros metabolitos secundarios y exoenzimas [82, 83]. Esta función se ve favorecida en contextos de alta densidad celular. Particularmente, en la bacteria *Pseudomonas protegens*, la síntesis de HCN brinda protección a las plantas de tabaco frente al mal de la podredumbre negra por prevenir la presencia de *Thielaviopsis basicola*, un patógeno fúngico de suelo [82]. Se ha propuesto que esta especie de bacterias beneficia a las plantas cuyas raíces habita, lo que indica que la síntesis de sus metabolitos secundarios puede utilizarse como factor natural de biocontrol de cultivos [65]. Así, el uso de bacterias productoras de HCN como bioplaguicidas ofrece un enfoque ecológico para una agricultura sostenible.

La síntesis de HCN está asociada a la familia de genes *hcn*. Esta familia está representada por el paquete de genes *hcnABC*, donde los subgenes A, B y C codifican para cada una de las unidades de la enzima HCN sintasa [84]. En particular, el ARNm de *hcnA* codifica la subunidad A de ese complejo trimérico, y es el mejor caracterizado de la familia. En parte, esto se debe a que en el año 2007

fue reportada la estructura tridimensional del complejo formado entre RsmE y dos segmentos de *hcnA* [26], fijados a cada uno de los sitios de unión de la proteína. Así, "$hcnA_{20-mer}$" es un constructo de 20 nucléotidos de longitud, formado por una secuencia de 12 bases provenientes de *hcnA* y 4 pares de bases flanqueantes que estabilizan su conformación para posibilitar su elucidación estructural. La secuencia de 12 nucleótidos contiene al motivo A(X)GGAX más próximo al codón de inicio de la traducción en la UTR 5' de *hcnA*. Trabajos posteriores sugieren que este es el primer motivo de la UTR de *hcnA* en unirse a RsmE [72].

La estructura reportada en la Referencia 26 muestra que el motivo A(X)GGAX, que en este caso adopta la secuencia específica ACGGAU, se encuentra en el extremo de un *stem-loop* al unirse a RsmE. Esto promueve que los 3 nucleótidos a un lado y otro del motivo ACGGAU se encuentren formando pares de bases. La estructura obtenida muestra que $hcnA_{20-mer}$ se une a una superficie altamente cargada positivamente de RsmE, formada por las láminas $\beta 1_A$, $\beta 4_B$ y $\beta 5_B$ en una de las caras de la proteína, así como a las láminas $\beta 1_B$, $\beta 4_A$ y $\beta 5_A$ en la cara contraria. Participan también de dicha interacción los *loops* entre $\beta 3/\beta 4$ y $\beta 4/\beta 5$ de cada sitio. Por lo tanto, RsmE interactúa con todos los nucleótidos tanto del *stem-loop* como los pares de bases C7-G14 y U6-A15. Esta característica es consistente con las demás estructuras reportadas para los ARNp que se unen a RsmE [72, 73]. La Figura I.5 muestra una representación pictórica para la secuencia de unión entre RsmE y $hcnA_{20-mer}$, así como del complejo final reportado por la Referencia 26.

Estudios complementarios indican que la unión entre RsmE y $hcnA_{20-mer}$ ocurre con un mecanismo alostérico de cooperatividad negativa [79]. Se postula que este fenómeno puede deberse a la distorsión del segundo sitio de unión de RsmE como consecuencia de la unión del primer oligómero. Esta hipótesis fue planteada en un trabajo que utilizó simulaciones del tipo MM-GBSA (por sus siglas *Molecular Mechanics - Generalized Born Surface Area*) para estimar la contribución de diferentes residuos a la Energía Libre de unión. Estas mismas observaciones se desprenden también de las constantes de afinidad para cada paso de unión, reportadas en la Referencia 72. Tanto el mecanismo de cooperatividad negativa como la contribución de nucleótidos individuales a la energía de unión entre RsmE y $hcnA_{20-mer}$, son

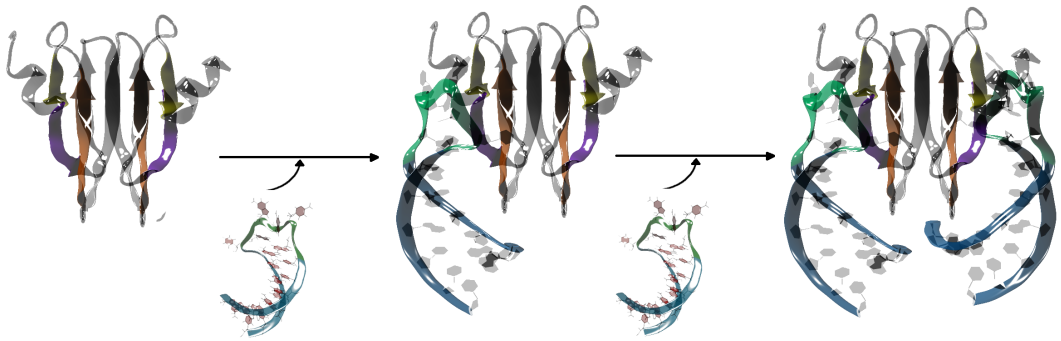


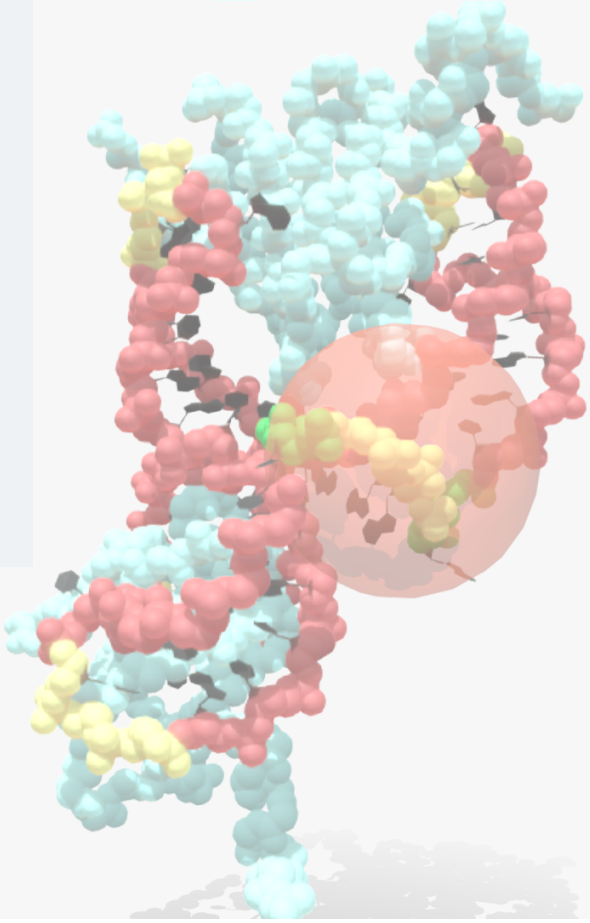
Figura I.5: Representación pictórica del orden seguido por $hcnA_{20-mer}$ al unirse a RsmE. Se destaca en verde al motivo ACGGAU de $hcnA_{20-mer}$, mientras el resto de su secuencia está representada en azul. Las láminas β de RsmE resaltadas son las que participan de la unión directa de $hcnA_{20-mer}$.

dos fenómenos retomados con enfoques alternativos en esta tesis (Ver Capítulo VI).

En este capítulo se presentó al sistema investigado en esta tesis. El mismo tiene a la proteína RsmE actuando como pivote central entre dos moléculas de ARN. Por un lado, se encuentra el ARNm sobre el que RsmE ejerce una función regulatoria a través de la unión a su segmento UTR 5'. Del otro lado se encuentra RsmZ, un ARNp capaz de evitar dicha interacción al unirse competitivamente a RsmE. La idea central de la tesis es que las características dinámicas, tanto de RsmE como de los ARN estudiados, determinan el mecanismo de su unión. Para arrojar luz sobre estas características, empleamos técnicas de DM. Estas técnicas fueron utilizadas con éxito, durante muchos años, en el estudio de la dinámica de proteínas. Sin embargo, su aplicación a los ARN es más reciente y tiene, por lo tanto, un grado de desarrollo menor. En virtud de lo expuesto, en los siguientes capítulos se discuten las propiedades estructurales y dinámicas de los ARN, así como las implicancias que las mismas tienen en las simulaciones de DM.

Espacio para anotaciones

II.1 Características estructurales de los ARN 22
II.1.1 Estructuras secundarias de los ARN . . 25
II.1.2 Estructuras terciarias de los ARN 31
II.1.3 Modularidad de ensambles y jerarquías estructurales de los ARN 33
II.2 Ejemplos de estudios computacionales sobre el ARN 41
II.2.1 Antecedentes sobre ARN pequeños (ARNp) 43



II. Estructura y dinámica de los ARN

Parte del desarrollo sobre los temas presentados en este Capítulo fueron incluidos en el libro "Bacterial Regulatory RNA: Methods and Protocols" de la serie "Methods in Molecular Biology" en el año 2023 [1].

Resumen de este capítulo Los ácidos ribonucleicos (ARN) poseen características dinámicas y estructurales que los diferencian marcadamente del resto de las biomoléculas. El reconocimiento de estos atributos es relativamente reciente, por lo que su estudio es actualmente un campo en constante crecimiento. Para ilustrar este punto, cabe mencionar que los primeros trabajos que reportaron estructuras cristalográficas del ARN datan de los años 70', casi dos décadas después de publicarse la primera estructura de una proteína. La gran diversidad estructural de estas moléculas desafía tanto los abordajes experimentales como computacionales, dado que es necesario observar al ARN con una mirada que difiere de la habitualmente utilizada para el estudio de proteínas globulares, haciendo hincapié en que su comportamiento está determinado por un ensamble de posibles conformaciones. Esta es, precisamente, el área de incumbencia

sobre la cual las técnicas computacionales pueden realizar su aporte. En este capítulo se introducen los conceptos básicos sobre estructura y dinámica de las moléculas de ARN, describiendo sus motivos estructurales secundarios y terciarios más recurrentes. A su vez se desarrolla el concepto de "modularidad de ensambles" (por su nombre en inglés "*ensemble modularity*"), enfatizando cuáles de los procesos involucrados en sus distintos niveles de jerarquía pueden ser observados por técnicas de DM. El capítulo finaliza mencionando algunos reportes previos de simulaciones computacionales abocadas al estudio del ARN, y sus limitaciones.

Los ARN poseen propiedades que los distinguen de otras moléculas más ampliamente estudiadas en el campo de la biofísica y la biología computacional. Estas características implican desafíos tanto para quienes abordan su estudio con técnicas experimentales como teóricas. En buena medida, esos desafíos se deben a la enorme flexibilidad de estas macromoléculas, en virtud de la cual es necesario dilucidar un conjunto de estructuras representativas y sus respectivas ponderaciones en lugar de una única estructura predominante. A nivel teórico, debido a que los tiempos característicos de sus cambios estructurales abarcan rangos mayores que en las moléculas más comúnmente estudiadas en el área (como es el caso de las proteínas globulares), es necesario adaptar las técnicas de análisis desarrolladas para proteínas al contexto particular de los ARN.

A continuación, se abordarán aspectos básicos sobre las estructuras de los ARN, para luego introducir los conceptos de "modularidad de ensambles" y "jerarquías estructurales".

II.1 Características estructurales de los ARN

El conocimiento que actualmente se dispone sobre las estructuras de los ARN está mayormente basado en la información obtenida por métodos experimentales, siendo la cristalografía de rayos X y RMN las técnicas a partir de las cuales se obtuvo la mayor parte de este conocimiento. En los últimos años, se produjeron importantes avances en el uso de Cryo-EM para la determinación estructural de moléculas de ARN libres de proteínas [85]. Esto necesariamente implica un sesgo en la información estructural disponible ya que sobre-representa a aquellos ARN que se

pliegan de manera funcional y bien definida, en desmedro de los más flexibles [86]. No obstante, muchos ARN no pertenecen a este último grupo. En esos casos, resulta esencial caracterizar su diversidad conformacional para comprender cómo llevan a cabo su función [87]. Por ello se hace necesario complementar la información obtenida por métodos experimentales con el aporte de estudios computacionales.

La primera estructura de un ARN que pudo ser resuelta data del año 1974, y corresponde al ARN de transferencia (ARNt) de la levadura, del aminoácido Alanina [88]. Es necesario destacar que este hito se hizo esperar nueve años tras haberse logrado el aislamiento de la molécula y la determinación de su estructura primaria (es decir, de su secuencia oligonucleotídica) [89]. Más aún, ocurrió dieciséis años después de haberse determinado la primera estructura cristalográfica de una proteína [90]. Esto retrata las dificultades propias de los estudios estructurales del ARN, lo cual impacta directamente sobre las simulaciones computacionales de estas biomoléculas debido a su efecto sobre los Campos de Fuerza (CF) empleados para dichas simulaciones [91] (Ver Sección III.3.1).

El primer aislamiento y posterior determinación estructural del ARNt permitió romper con las nociones previas que circunscribían la función biológica de los ARN a ser meros intermediarios entre el almacenamiento de la información genética y su expresión [92]. Esto inauguró una nueva etapa en la comprensión sobre los ARN que impulsó un gran número de estudios, destacándose aquellos que profundizaron sobre la existencia de ARN auto-plegados capaces de catalizar reacciones biológicas y acelerarlas en magnitudes comparables con las enzimas proteicas [93, 94]. A este tipo de moléculas se las denominó ribozimas, y su descubrimiento permitió no sólo expandir el Dogma Central de la Biología Molecular, sino también reforzar la hipótesis del "mundo de ARN", la cual establece que que los sistemas basados en esta molécula fueron las primeras formas de vida [95–97].

El ribosoma, molécula responsable de la traducción del ARNm a proteínas, puede considerarse de hecho un caso particular de una ribozima [98]. Dicha molécula, de relevancia para el sistema abordado en esta tesis, fue objeto de una gran cantidad de estudios abocados a la determinación de su estructura tridimensional. La información así obtenida permitió incrementar enormemente los conocimientos

actuales sobre los ARN en general, ya que buena parte de los motivos estructurales que hoy se entienden como típicos de estas moléculas surgen a partir de estudios sobre el ribosoma [86].

Para comenzar a hablar de los motivos típicos de los ARN, los cuales están fundamentalmente dictados por sus estructuras secundarias y terciarias, es necesario mencionar en primer lugar a su estructura primaria. La misma está definida, como ocurre con las proteínas, por la secuencia (en este caso, oligonucleotídica). En la naturaleza, la secuencia de los ARN está definida por la combinación de los nucleótidos Adenina (A), Uracilo (U), Citosina (C) y Guanina (G) [89]. Cada una de estas bases definen la identidad de su nucleósido correspondiente, el cual se completa con un anillo de ribosa unido a la base en su posición C1' y un grupo fosfato unido a su carbono C4'. Así, las cadenas de ARN se construyen a través de enlaces fosfodiéster entre estos residuos azucarados, de manera que cada carbono ubicado en el extremo 3' se une al carbono del extremo 5' anterior. En consecuencia, las cadenas de ARN son asimétricas, por lo cual poseen un sentido de lectura. El extremo 5' se considera el inicio por ser la posición en la que comienza su síntesis en los organismos vivos [86]. La figura II.1 resume los conceptos introducidos en este párrafo.

La estructura primaria tiene una incidencia crucial para el ordenamiento del ARN en su estructura secundaria, la cual queda determinada por el apareamiento de bases durante el plegamiento de la molécula. La mayoría de estos apareamientos se rigen por el patrón canónico de Watson-Crick, en el que A interactúa con U, y C con G [95]. Estos pares son particularmente estables y poseen energías de interacción asociadas que oscilan entre 1 a 3 $kcal/mol$ dependiendo del contexto estructural [99–101]. Otro tipo relevante de apareamiento entre bases son las interacciones de Hoogsteen, las cuales utilizan la cara C6–N7 de las purinas (es decir, de los residuos A y G) para formar puentes de hidrógeno con la cara N3-C4 de las pirimidinas (que incluye a los residuos U y C) [102]. Los apareamientos inversos de Hoogsteen involucran un giro de 180° de una de las bases en relación a su apareamiento original. La figura II.2A compara los apareamientos del tipo Watson-Crick con las interacciones de Hoogsteen. Otra interacción frecuentemente encontrada, y

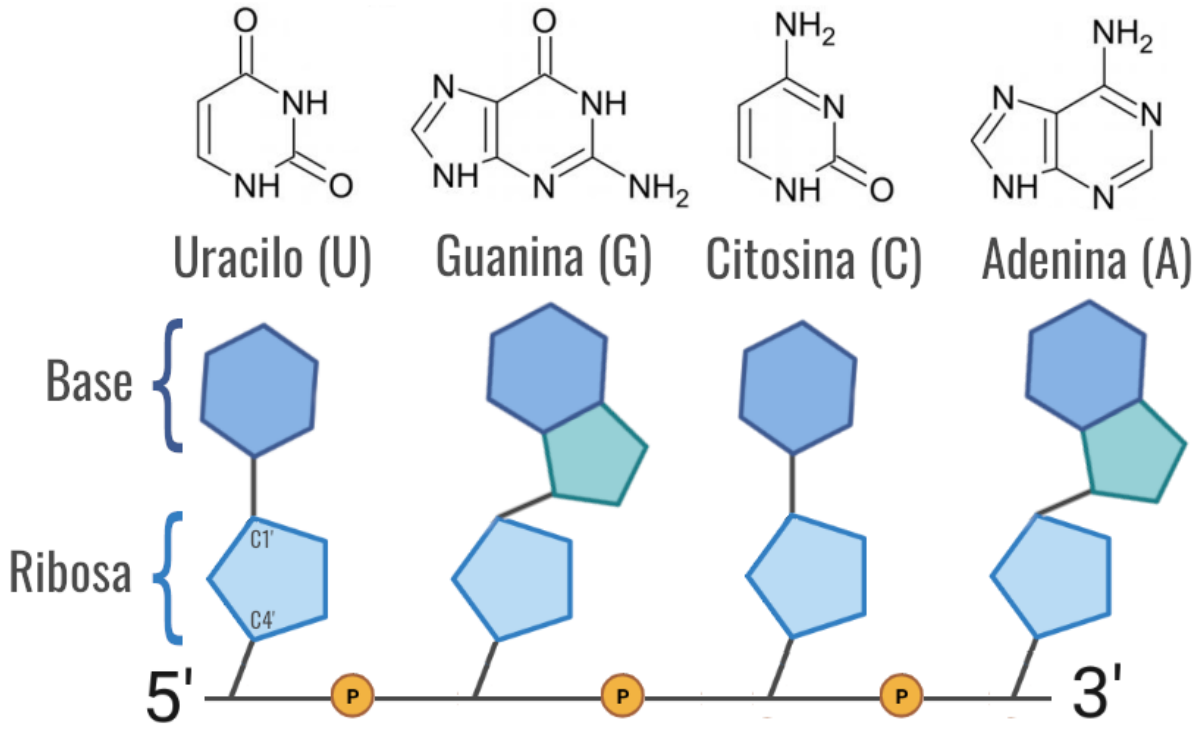


Figura II.1: Representación pictórica de las bases que determinan la estructura primaria del ARN y de la manera en la que los nucleótidos se ensamblan entre sí a través de enlaces fosfodiéster. Se distingue en cada nucleósido de la cadena a su base unida a la ribosa, destacando al carbono interviniente en dicha unión.

relevante por sus implicancias en la estructura de los ARN, es la interacción G/U [103].

II.1.1 Estructuras secundarias de los ARN

La estructura secundaria de los ARN está formada por arreglos ordenados pequeños, sostenidos por interacciones canónicas y no canónicas entre pares de bases. Una interesante consecuencia de lo dicho es que los motivos estructurales típicos de los ARN están sujetos a restricciones basadas en secuencias consenso, de manera que cambios en un nucleótido dentro del motivo suelen estar acompañados por modificaciones de su región complementaria para preservar la estructura del motivo y, por ende, su función [104–111]. Este aspecto, también conocido como "isoestericidad",

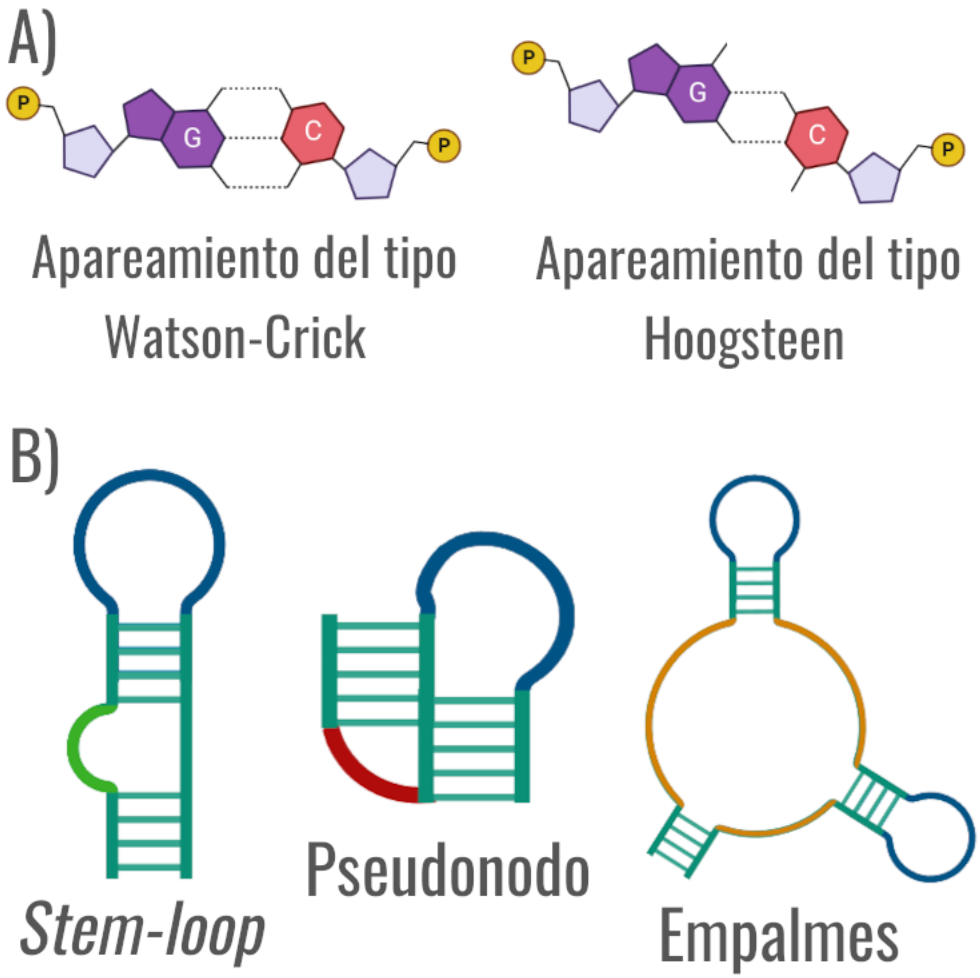


Figura II.2: Representación pictórica de apareamientos y estructuras secundarias típicas de los ARN. **Panel A**: Representación de los apareamientos del tipo Watson-Crick y Hoogsteen entre una Guanosina (color violeta) y una Citosina (color naranja). Los puentes de Hidrógeno entre ambas moléculas se representan con líneas punteadas **Panel B**: Representación pictórica para las estructuras secundarias de un *stem-loop*, un pseudonodo, y un empalme de tres vías. Las regiones apareadas se representan en verde, mientras las regiones de simple cadena correspondientes a *loops*, burbujas, uniones de pseudonodos y empalmes se representan en azul, verde claro, rojo y amarillo, respectivamente.

es fundamental para comprender el desarrollo evolutivo de los ARN y sus patrones de covariancia.

Algunas de las principales características de los motivos estructurales de los ARN se resumen a continuación [86]:

- **Son unidades modulares, discretas, e indivisibles**. Esto implica que afectar alguna de sus interacciones de puente de Hidrógeno individuales o de apilamiento generalmente causa la disrupción del motivo completo.
- **Normalmente son autónomos,** es decir, el mismo plegamiento o motivo puede ocurrir en diferentes entornos moleculares porque están dictados casi exclusivamente por su secuencia.
- **Son de carácter recurrente,** es decir, que son transversales a la totalidad de los ARN independientemente de su historial evolutivo.
- **Tienen múltiples y diversos roles biológicos.** Esto implica que si bien el plegamiento ocurre independientemente de su contexto estructural, éste último influye sobre su comportamiento una vez plegado.

Las estructuras secundarias más típicas se encuentran representadas pictóricamente en la Figura II.2. A continuación se describe a cada una de ellas.

Hélices dobles y *stem-loops*

Las hélices dobles son uno de los motivos más conocidos entre las moléculas de ARN y tienen el rol de proveer estabilidad estructural a la molécula. Aquellas dictadas por apareamientos canónicos suelen ser cortas, con no más de 12 pares de bases consecutivos [86]. Para ilustrar su importancia, basta con mencionar que estos motivos contribuyen a la estabilidad estructural de los ARN aún más que las estructuras secundarias análogas de las proteínas [99–101]. Las hélices muchas veces presentan una alternancia entre regiones de nucleótidos apareados y no apareados. A las regiones no apareadas comúnmente se las conoce como "*loops*" y en ciertas ocasiones son capaces de formar pares no canónicos [112]. Por simplicidad, me referiré a partir de ahora al motivo estructural completo, formado por regiones apareadas en forma de hélice que finalizan en un segmento de simple cadena, como "*stem-loops*". Los *stem-loops* estabilizan otros motivos dentro de la misma molécula a través de las interacciones que son capaces de establecer (ver Sección II.1.3).

Más allá de su rol estructural, las hélices dobles y los *stem-loops* cumplen

también funciones relacionadas con la acción biológica de la molécula. Un ejemplo de particular relevancia para este trabajo es la participación de los *stem-loops* en la interacción con otras macromoléculas. Estudios recientes sugieren que los *stem-loops* de estructura estable están casi siempre asociados a algún tipo de interacción con proteínas [113]. Dichas interacciones están generalmente, pero no exclusivamente, mediadas por los segmentos de simple cadena en los extremos de los *stem-loops*.

Otro de los procesos comúnmente regulados por interacciones mediadas por *stem-loops* es, precisamente, la traducción del ARNm [114]. A su vez, se ha reportado que el efecto que tienen estos motivos sobre la tasa traduccional del ribosoma fluctúa dependiendo de la extensión del *stem-loop*, la presencia de burbujas (ver Figura II.2B), y su composición GC (es decir, de pares de base entre Guaninas y Citosinas) [114, 115]. En contrapartida, existen ejemplos de moléculas de ARN que, al interactuar con el ribosoma a través de uno de sus *stem-loops*, favorecen la traducción de un determinado gen [116]. En el Capítulo I se describió un sistema donde el ARNp RsmZ es capaz de secuestrar a la proteína RsmE, inhibiendo la acción represora de esta última sobre la traducción del gen *hcnA*. Dicho mecanismo se produce, precisamente, a través de las interacciones formadas entre la proteína y los *stem-loops* tanto de los ARNp que la regulan como del gen en cuestión [73].

Pseudonudos

Los pseudonudos (en inglés *pseudoknots*) son motivos estructurales muy recurrentes en moléculas de ARN. El primer reporte de este motivo data del año 1982, cuando es detectado en la estructura del ARNt del patógeno de plantas conocido como virus del mosaico amarillo [117]. Este primer reporte se constituyó en una referencia a partir de la cual se pudo detectar el mismo motivo en un mayor número de moléculas, estableciéndose como un principio general del plegamiento del ARN [86].

Los pseudonudos son estructuras que se componen mínimamente de dos hélices conectadas por regiones de simple cadena (Ver Figura II.2B). Si bien existen diversas topologías asociadas a este motivo, la más ampliamente extendida y mejor

caracterizada es la de tipo "H", que se forma cuando las bases del segmento de simple cadena de un *stem-loop* forman pares intramoleculares con bases fuera del tallo [118]. Ese caso es, precisamente, el que se muestra en la Figura II.2B. Como consecuencia de esas interacciones, se forma un segundo *stem-loop*, dando como resultado un pseudonudo con dos hélices y dos regiones de simple cadena [119]. Es importante destacar que el tamaño de los *stem-loops* implicados en la formación de un pseudonudo altera las características estructurales de los mismos. A modo de ejemplo, las regiones helicoidales de los *stem-loops* que forman un pseudonudo pueden establecer interacciones con otras regiones helicoidales, formando ramificaciones de pseudonudos [120].

El contacto con otros pseudonudos es una de las interacciones más frecuentes de este motivo [118]. Dicha interacción está mediada por los segmentos de simple cadena de hélices adyacentes, las cuales forman puentes de hidrógeno que estabilizan al motivo. Esto implica que una de las principales funciones biológicas asumidas por los pseudonodos sea del tipo estructural a nivel intramolecular. No obstante, ellos pueden asumir otras funciones biológicas asociadas a su interacción con otras moléculas.

Entre las funciones biológicas más relevantes de los pseudonodos, se destacan el formar parte del núcleo catalítico de varias ribozimas [121] y del propio ribosoma [122], participar en el auto-*splicing* de intrones [123], establecer interacciones fundamentales en el mecanismo de acción de la telomerasa [124, 125] y desempeñar un papel fundamental en la alteración de la expresión génica por inducción de cambios en el marco de lectura ribosomal en virus [118, 119, 126–128]. Recientemente, se reportó que un pseudonudo presente en el extremo 5' del genoma de SARS-CoV-2 afecta la replicación de su ARN, por provocar el corrimiento en su marco de lectura ribosomal [129].

Empalmes (*junctions*)

Los empalmes (en inglés, "*junctions*") son segmentos de simple cadena que unen motivos estructurales con forma de hélice [130]. Se distinguen de las cadenas simples presentes en *stem-loops* y pseudonudos porque no forman parte de una

estructura secundaria mayor, sino que son un motivo estructural en sí mismo. La figura II.2B muestra un ejemplo de empalme entre tres hélices (siendo esa la mínima cantidad de regiones de doble cadena unidas por este motivo para ser definido como tal [86]), donde dos de ellas pertenecen a su vez a *stem-loops*. La base de datos *RNAJunction* (disponible en: https://rnajunction.ncifcrf.gov/) recopila información de motivos estructurales de este tipo.

Debido a estar formados por segmentos de simple cadena relativamente largos, de longitud variable y generalmente asimétricos [131], los empalmes generalmente son regiones de alta flexibilidad dentro del ARN [132]. No obstante, se trata de estructuras bien definidas que contribuyen significativamente a la arquitectura de la molécula dentro de la cual están presentes [130]. Por ejemplo, los empalmes influyen sobre las conformaciones adoptadas por las hélices unidas a ellos, de manera que éstas tienden a adoptar patrones estructurales similares incluso cuando están aisladas del contexto molecular original [133]. Asimismo, los empalmes tienden a aproximar entre sí a las hélices que emergen de ellos, de manera que se ven favorecidas interacciones tales como los apilamientos coaxiales que forman una pseudo-hélice continua o motivos A-menores (ver Sección II.1.2) [134–137].

El criterio de clasificación más habitual de los empalmes es el número de vías que emanan de los mismos. Los tipos más ampliamente estudiados son los de tres y cuatro vías, que desempeñan funciones biológicas más allá de su rol estructural [131, 138]. Los empalmes de tres vías forman el centro catalítico de ribozimas, estabilizando su estructura y aumentando así la eficiencia de las mismas [139–141]. Por su parte, interacciones de apilamiento entre hélices unidas a empalmes de tres vías determinan las conformaciones holo y apo de ciertos *riboswitches* [142]. Tanto los empalmes de tres vías como los de cuatro están presentes en la arquitectura del ARN ribosomal (ARNr) 23S, comprendido dentro de la subunidad 50S [138]. Se ha reportado que los empalmes de cuatro vías aceleran el plegamiento de ribozimas inducido por iones [143] y, en determinados contextos, participan en la regulación traduccional [144].

El carácter polimórfico y dinámico de los empalmes de cuatro vías implica, a su vez, un alto nivel de diversidad estructural, con conformaciones generalmente

estabilizadas por interacciones terciarias adicionales, o bien por su unión a proteínas. Como se mencionó en el Capítulo I, el ARNp del sistema de regulación post-transcriptional RsmE-RsmZ posee un empalme de cuatro vías, que no sólo determina la conformación del ARN estudiado sino su capacidad de unión a la proteína sobre la que ejerce una función regulatoria.

II.1.2 Estructuras terciarias de los ARN

De la misma manera que la interacción entre los oligonucleótidos que forman la estructura primaria del ARN determina su estructura secundaria, la interacción entre los motivos secundarios define su estructura terciaria. Típicamente, dichas interacciones involucran puentes de Hidrógeno a través de grupos 2'-OH [86]. Los tiempos característicos para la ruptura y formación de estas interacciones las convierte en un blanco ideal para ser observadas en simulaciones de DM. Sobre este aspecto en particular se ahondará en la Sección II.1.3.

A continuación se mencionan algunas de las interacciones terciarias más típicas de los ARN. La Figura II.3 presenta, con una representación pictórica, a las estructuras terciarias aquí descritas.

- *Ribose zipper*: Este motivo se caracteriza por varias interacciones consecutivas con al menos dos puentes de hidrógeno base-esqueleto entre el átomo N3 de una purina o el átomo O2 de una pirimidina en el extremo 5', y el grupo 2'-OH de una ribosa en el extremo 3' del lado opuesto. Estos contactos terciarios también implican interacciones tanto entre bases, como de las bases con el esqueleto del ARN. Se ha reportado la existencia de al menos once tipos de *ribose zippers*, la mayoría de los cuales son observados en los ARNr [145]. Estos motivos típicamente implican interacciones entre segmentos de simple cadena y están habitualmente ligados a interacciones con proteínas. En la Sección IV.3.2 se discute cómo una interacción de este tipo podría ser relevante para la unión de RsmE a RsmZ.

- *Kissing loops*: Es una estructura formada por la interacción entre dos *stem-loops* a través de sus segmentos de simple cadena. En general implican el contacto entre regiones distantes en una molécula de ARN, y están mediadas

por nucleótidos que interactúan entre sí a través de pares de bases complementarios del tipo Watson-Crick [146]. Se ha encontrado este tipo de interacciones en la unidad ribosomal 23S de *E. coli*, así como en moléculas de ARNm y ARNt [112, 147], y está ampliamente estudiado su rol en el reconocimiento de proteínas [148]. Por otra parte, se ha descrito que los retrovirus utilizan interacciones de este tipo para dar inicio a la dimerización del ARN genómico. El virus del VIH presenta un ejemplo de este fenómeno [147]. Las interacciones entre regiones desapareadas de *stem-loops* se discuten en la Sección IV.3.2 por su importancia para definir estados accesibles u ocluidos de RsmZ a la hora de unir a RsmE.

- Motivo A-menor: Este motivo involucra a un nucleótido de Adenina insertado en el surco menor de una hélice de ARN. Es un motivo abundante y está involucrado en interacciones que estabilizan la estructura terciaria de los ARN. Existen cuatro tipos de motivos A-menor que se diferencian en las posiciones adoptadas por los átomos O2' y N3 del residuo de Adenina, respecto al átomo O2' de la base receptora de la hélice [86]. Estos motivos son comunes en interacciones con *stem-loops*, contactos con empalmes, y secciones donde ocurren cambios en la dirección del esqueleto del ARN. La estabilidad de los motivos A-menor, particularmente los del tipo I, está dada por la estrecha interacción entre la Adenina y los pares G-C dentro del surco menor de la hélice. El alto grado de conservación filogenética tanto de las Adeninas como de los pares G-C con los que interactúan, particularmente en el ARNr, dan cuenta de su importancia estructural y biológica [149]. Las Adeninas involucradas en motivos A-menor muchas veces se encuentran apiladas en lo que se conoce como "parches de Adeninas". Rara vez son más de tres los residuos de Adenina que forman este motivo y como regla general, las interacciones de este tipo disminuyen a lo largo de la dirección 5' a 3' [150].

Dada la marcada diversidad conformacional de los ARN en comparación con otras macromoléculas, resulta necesario caracterizar un conjunto amplio de estructuras representativas para comprender de qué manera ejercen su función biológica. En la siguiente sección se detallarán las características típicas de dicha diversi-

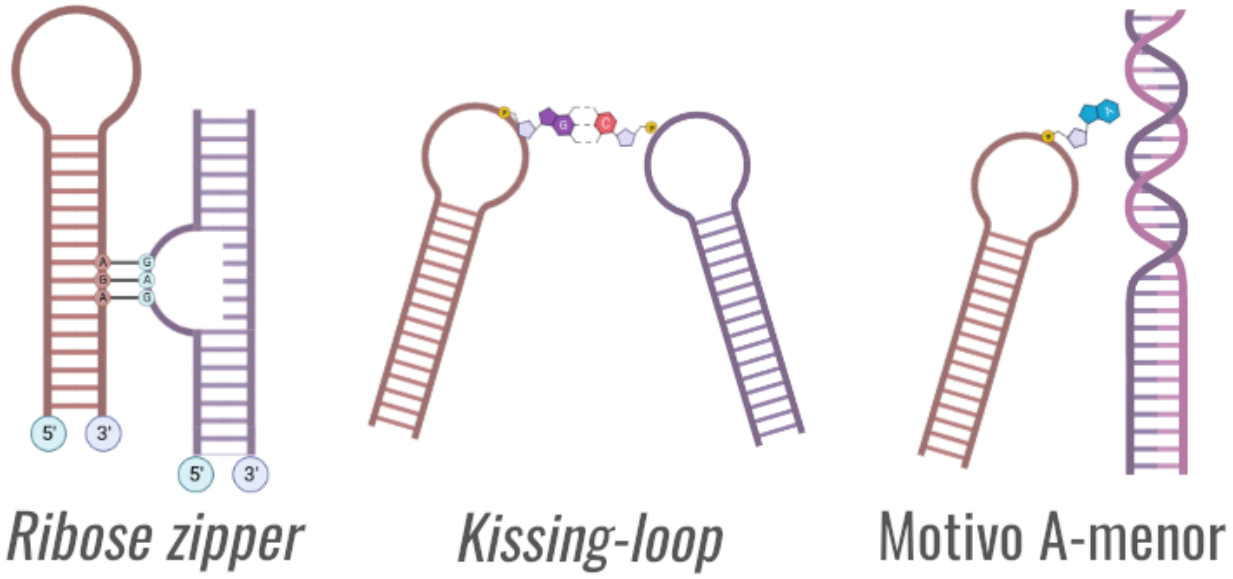


Figura II.3: Representación pictórica para las estructuras terciarias más comunes entre las moléculas de ARN. Así, se presenta un motivo del tipo *ribose zipper* entre un segmento helicoidal y un fragmento simple cadena antiparalela; un motivo del tipo *kissing loop* mediado por un apareamiento de Watson y Crick entre una Guanina (violeta) y una Citosina (naranja); y un motivo A-menor entre una Adenina (verde) y el surco menor de una hélice.

dad conformacional, así como de las escalas de tiempo en las que ocurren las transformaciones estructurales susceptibles a ser observadas por DM.

II.1.3 Modularidad de ensambles y jerarquías estructurales de los ARN

Un concepto fundamental para explicar/describir la diversidad conformacional de los ARN es el de las Superficies de Energía Libre (SEL). Este se desarrolló y resultó útil en los estudios de plegamiento y cambios conformacionales de las proteínas. Si bien existen diferencias marcadas entre el comportamiento de las proteínas plegables y los ARN, las mismas pueden ser racionalizadas por las diferencias en sus SEL.

A una cierta temperatura T, un sistema con N partículas que ocupan un volumen V y que está en equilibrio, tiene una energía libre definida $A(N,V,T)$. De acuerdo a

los fundamentos de la termodinámica estadística, para un sistema cuyo potencial no depende de las velocidades este número puede evaluarse mediante,

$$A(N,V,T) = -k_B T \ln\left(Z(N,V,T)\right) + C(T) \tag{II.1}$$

En esta expresión, $C(T)$ es una constante que depende de la temperatura, mientras que $Z(N,V,T)$ es la integral configuracional,

$$Z(N,V,T) = \iiint e^{-V(\mathbf{x})/k_B T} d\mathbf{x}, \tag{II.2}$$

donde $V(\mathbf{x})$ es la energía potencial de una determinada configuración $\mathbf{x}$ (la cual depende de múltiples variables), y la integral múltiple debe realizarse sobre todos los grados de libertad del sistema contenidos en II.2.

Una vez fijados N, V y T, la Energía Libre de un sistema en equilibrio es un número. No obstante, en el campo de la Química, Bioquímica y Biofísica computacional se ha hecho una extensión de la expresión II.1 que consiste en obtener una "Energía Libre" en función de una, o unas pocas coordenadas cuidadosamente elegidas.

Si estas coordenadas estuvieran entre las variables de integración de la ecuación II.1, el cálculo consistiría simplemente en realizar todas las integrales requeridas en II.2 excepto las que involucran a las variables seleccionadas. Sin embargo, esta situación es atípica ya que generalmente las variables relevantes dependen de manera relativamente compleja de las variables de integración. Por este motivo es necesario introducir dentro de la integral a la función de delta de Dirac. Esta función proyecta la integral sobre la o las coordenadas seleccionadas. Por ejemplo, para dos coordenadas generalizadas ξ_1 y ξ_2, la integral a evaluar es,

$$Z(N,V,T,\xi_1^*,\xi_2^*) = \iiint \delta(\xi_1^* - \xi_1(\mathbf{x}))\delta(\xi_2^* - \xi_2(\mathbf{x}))e^{-V(\mathbf{x})/k_B T} d\mathbf{x}. \tag{II.3}$$

Utilizando una integral configuracional como la de la Ecuación II.3 en la Ecuación II.1, se evalúa una SEL $A(N,V,T,\xi_1,\xi_2)$. En cambio, si solo se proyecta

sobre una coordenada el resultado es un Perfil de Energía Libre (PEL), también llamado Potencial de Fuerza Media (PMF, por sus siglas en inglés), mientras que si se proyecta sobre más de dos variables se obtiene una hipersuperficie. La Figura II.4 muestra una representación pictórica de esta clase de superficies.

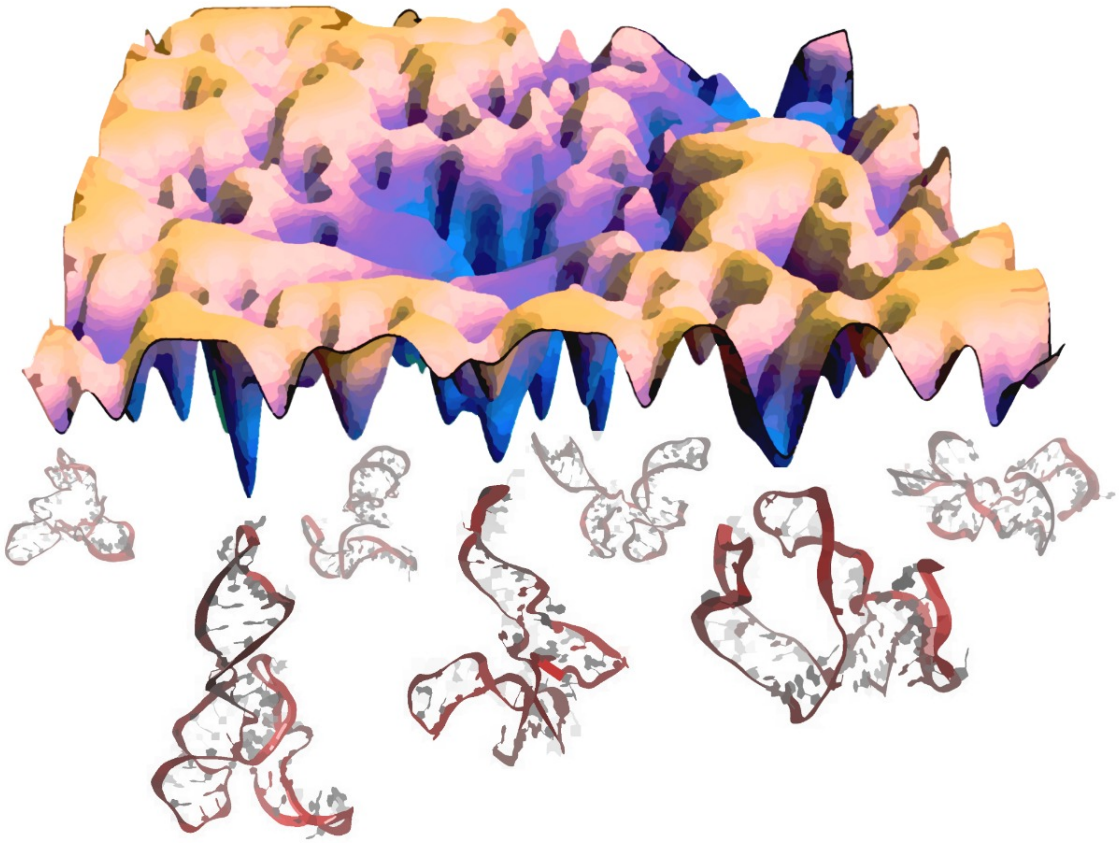


Figura II.4: Representación pictórica de una SEL para una molécula de ARN. Los mínimos de la superficie corresponden a las estructuras que constituyen un mismo conjunto conformacional. Imagen adaptada de la Referencia 1.

La elección de las coordenadas sobre las cuales conviene calcular la SEL debe evaluarse para cada caso en particular. A veces, la elección puede realizarse en base al conocimiento de lo que ocurre en el proceso que se quiere estudiar. Por ejemplo, en este trabajo de tesis, se estudió el proceso de disociación de un dímero de RsmE desde un complejo formado entre RsmZ y RsmE. En tal caso, resulta claramente conveniente definir la coordenada de reacción como la distancia entre dos puntos, uno de los cuales es el centro de masa de una región de RsmE y el otro es el centro de masa de una región de RsmZ. En otras ocasiones, la elección no es tan obvia y es necesario recurrir a herramientas más sofisticadas para la definición.

A modo de ejemplo, en la Sección III.5.2 se describe cómo los primeros componentes principales de un sistema pueden utilizarse como coordenadas para evaluar sus SEL. Por el momento, basta con decir que estas coordenadas describen las fluctuaciones de mayor amplitud del sistema y que, por lo tanto, son apropiadas para identificar los cambios conformacionales más significativos captados por la simulación. Recientemente, se han reportado diversos algoritmos que hacen uso de la Inteligencia Artificial (IA) para definir las coordenadas colectivas óptimas que describen un dado proceso [151]. Si bien estas propuestas parecen prometedoras en este momento, su verdadera utilidad va a ser demostrada con el paso del tiempo.

En cualquier caso, una vez que se identificaron las variables relevantes $\{\xi_i\}$ y se calculó la SEL en función de las mismas, $A(N,V,T,\{\xi_i\})$, la función obtenida permite comprender diversos aspectos del funcionamiento del sistema. Para ello, es necesario notar la relación que existe entre esta función y la probabilidad de hallar una configuración determinada. La misma está dada por,

$$A(\{\xi_i\}) = -k_B T \ln\left(\rho(\{\xi_i\})\right), \tag{II.4}$$

donde deliberadamente se omitió explicitar la dependencia de A con N, V y T para simplificar y para destacar su relación con la Función de Densidad de Probabilidad (FDP) $\rho(\{\xi_i\})$ de las variables seleccionadas. En virtud de la Ecuación II.4, se observa que las regiones de alta probabilidad del espacio configuracional, correspondientes a estructuras estables o metaestables, corresponden a mínimos en la SEL mientras que las estructuras más improbables corresponden a los máximos.

De esta manera, las barreras ubicadas entre mínimos adyacentes determinan la dificultad para pasar de uno al otro. En resumen, el conocimiento de la SEL del sistema provee la información necesaria para comprender tanto su comportamiento termodinámico como cinético.

La estabilidad relativa de dos conformaciones X y Y es medida por la diferencia de Energía Libre entre las mismas,

$$A(\{\xi_i\}^Y) - A(\{\xi_i\}^X) = -k_B T \ln\left(\frac{\rho(\{\xi_i\}^Y)}{\rho(\{\xi_i\}^X)}\right), \tag{II.5}$$

Así, la velocidad para pasar de un pozo a otro depende de la barrera de Energía Libre que los separa de acuerdo con una relación tipo Arrhenius.

Las consideraciones realizadas hasta este punto permiten describir las jerarquías estructurales de los ARN [152], las cuales deben su existencia a la rugosidad de las SEL de estas moléculas. Esta rugosidad abarca tres escalas diferentes, como se ilustra en la Figura. II.5. Existe un nivel más básico, también llamado Nivel 1, en el que las diferentes conformaciones estables están separadas por grandes barreras de energía. Estas conformaciones son, por ende, muy estables. Al amplificar la vista sobre cualquiera de los pozos de este primer nivel, se observa que el mismo también es rugoso, con barreras de Energía Libre menores que las del Nivel 1. Por lo tanto, la estabilidad de las diferentes conformaciones del Nivel 2 es menor que la del 1. Finalmente, haciendo *zoom* sobre alguno de los pozos del Nivel 2, también se observa que el mismo es rugoso, con barreras de Energía Libre aun más bajas que las del nivel anterior. Por este motivo, interconversiones entre los pozos del nivel 3 son las más frecuentes de todas.

En los ARN, el primer nivel jerárquico corresponde a los motivos estructurales que están involucrados en la arquitectura secundaria. Por lo tanto, la interconversión entre conformaciones de este nivel implica la alteración tanto de estructuras terciarias como secundarias, con altos costos de energía y tiempos largos de transición que van desde los milisegundos hasta varios segundos, e incluso horas, dependiendo

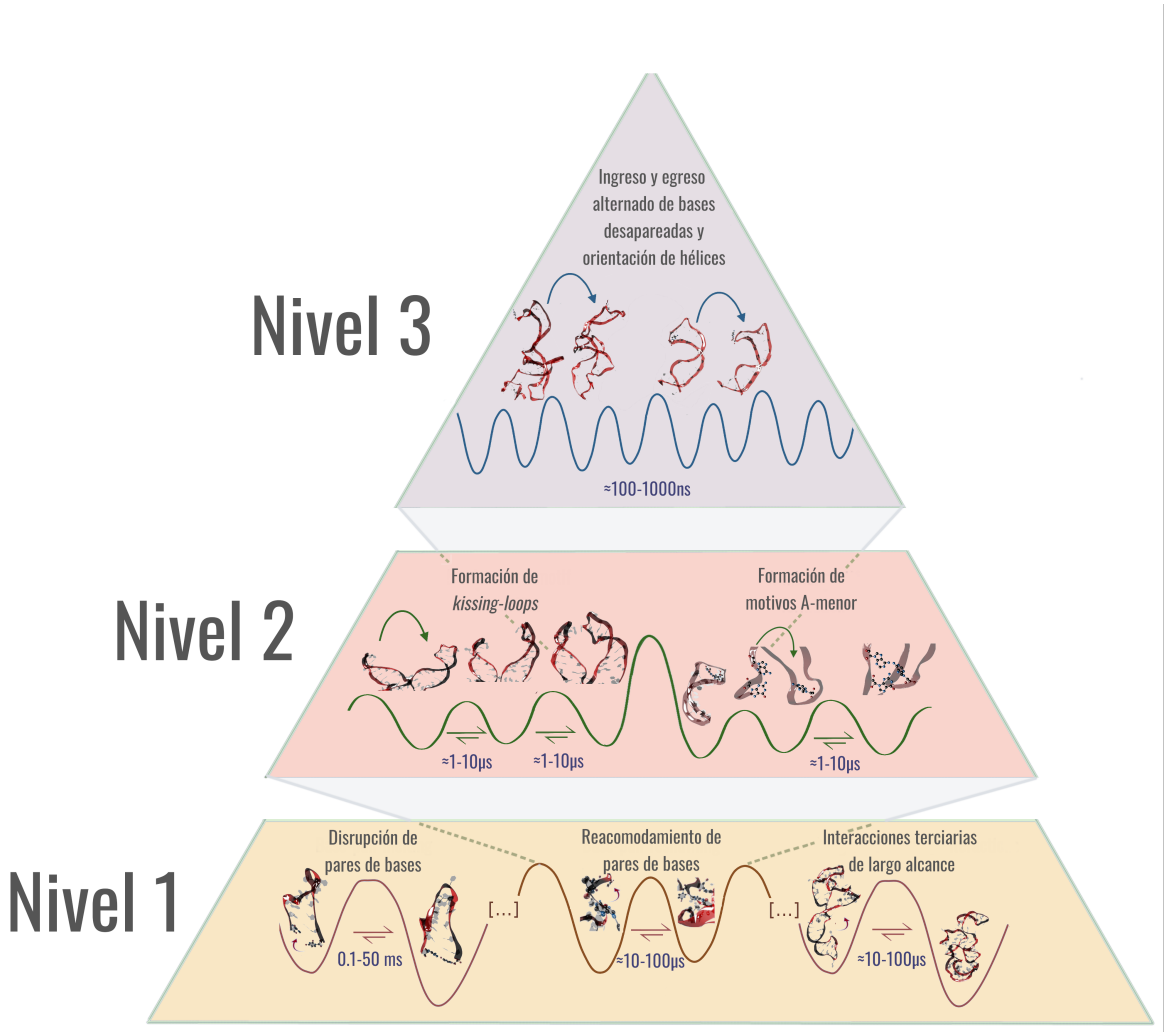


Figura II.5: Representación pictórica para la jerarquización de un ensamble conformacional y de los tiempos involucrados para la transición entre las conformaciones de cada nivel. Imagen adaptada de la Referencia [1].

.

de las interacciones interrumpidas [152, 153]. Algunos ejemplos de procesos que se encuentran este nivel son la disrupción de pares de bases, el reacomodamiento e isomerización entre estas y los cambios conformacionales que involucran desplazamientos de gran amplitud mediados por regiones flexibles. Estas interacciones usualmente involucran la exposición alternada de diferentes motivos funcionales,

de manera que están implicadas en procesos de reconocimiento a otras moléculas [152].

Para ilustrar los tiempos característicos de las transiciones en este primer nivel, cabe mencionar que la disrupción de pares de bases ocurre en escalas que van desde los 0.1 a 50 ms, dependiendo de la naturaleza de los residuos involucrados en las interacciones de apilamiento. En el caso de reacciones de catálisis mediadas por moléculas de ARN, ocurren cambios estructurales globales en escalas que comprenden desde los nanosegundos a los minutos, para posicionar adecuadamente a los residuos participantes de la reacción en el núcleo catalítico [154]. Muchos de estos tiempos están fuera de lo permitido para un normal funcionamiento de la homeostasis en organismos vivos. Por este motivo, existen proteínas del tipo chaperonas abocadas al correcto plegado del ARN, que aceleran los cambios en la arquitectura de los mismos en respuesta a determinados estímulos. Así, se reducen los tiempos implicados hasta el rango de los milisegundos [155, 156].

Dado que las simulaciones de DM para las moléculas de ARN alcanzan al día de hoy escalas de microsegundos (excepcionalmente, de milisegundos) [157, 158], el estudio de los eventos comprendidos dentro de este nivel de jerarquización está mayormente fuera de su alcance. No obstante, algunos eventos en particular, como el rearreglo de apilamientos en el par G-A o la isomerización de bases, pueden ocurrir más frecuentemente que las transiciones de estructuras secundarias, siendo más factible su observación en simulaciones de DM [153]. Por otra parte, existen adaptaciones de la DM estándar que permiten explorar transiciones entre estados separados por altas barreras de energía libre de la SEL. Estas se discutirán con más detalle en la Sección III.4.

El segundo nivel de jerarquía en las estructuras de los ARN comprende transiciones entre conformaciones con diferencias menores desde el punto de vista de sus estructuras secundarias. Dichas transformaciones típicamente involucran la ruptura de pares de bases individuales que no requieren de la asistencia de chaperonas [146]. Se incluye también en este nivel a las transiciones alternadas entre apareamientos de bases ubicados en *stem-loops*, la ruptura y formación de pares de bases no canónicos, y las alteraciones de la estructura terciaria que no involucran grandes

desplazamientos.

Ejemplos típicos de las interacciones comprendidas en este nivel son los *kissing-loops*, los apilamientos de bases entre motivos estructurales cercanos, unión a cationes, tripletes de bases y los motivos A-menores [146]. Estas interacciones son biológicamente relevantes para la estabilización de la estructura tridimensional del ARN, dado que permiten la oclusión o exposición de residuos involucrados en el reconocimiento de otras moléculas o la alternancia entre conformaciones activas e inactivas [152]. En particular, la unión de proteínas a segmentos de simple cadena en los extremos de los *stem-loops* suelen estabilizar conformaciones que favorecen la unión de una segunda proteína. Así, los contactos mediados por motivos de este nivel a menudo se relacionan con procesos de unión cooperativos [159]. El complejo RsmZ-RsmE es un ejemplo de este fenómeno [73], como se describió en el Capítulo I y se profundizará en el Capítulo IV.

Dado que este nivel de organización implica diferencias más sutiles respecto al primero en cuanto a la arquitectura secundaria del ARN, presenta una mayor preponderancia de conformaciones iso-energéticas, es decir, igualmente estables [152]. Cuando las barreras que separan a dichos estados son lo suficientemente bajas, la interconversión entre estructuras es más rápida que en el primer nivel, quedando en el rango de los microsegundos a los milisegundos [153]. Por lo tanto, algunas interacciones de este nivel se acercan al límite superior alcanzado por técnicas de DM clásicas.

En el tercer nivel jerárquico se encuentran las transiciones entre conformaciones que tienen estructuras secundarias iguales, o muy similares, pero que difieren principalmente en cómo estas se disponen respecto al resto de la molécula. O sea, los cambios conformacionales de este nivel están asociados a movimientos e interacciones menos dependientes de la secuencia y más fuertemente ligadas a su estructura tridimensional [133, 160].

Las escalas de tiempo de este nivel van desde los picosegundos a los microsegundos, por lo que son susceptibles de ser observadas en simulaciones de DM. Algunos ejemplos de eventos de este nivel son la alternancia en la exposición de regiones de simple cadena, el reacomodamiento de bases desapareadas, las reorientaciones

de fosfatos, y los movimientos colectivos de hélices [152], los cuáles ocurren en escalas de tiempo comprendidas entre los nanosegundos a los microsegundos [161–163]. La excepción a esta tendencia general ocurre con los empalmes de cuatro vías, dado que los mismos presentan transiciones en escalas de los milisegundos al segundo como consecuencia de las interacciones de apilamiento cooperativo [138].

Por último, teniendo en cuenta las jerarquías estructurales de los ARN, cabe discutir el concepto de "modularidad de ensambles". El mismo tiene su origen en numerosos estudios que han firmemente establecido que las estructuras de los ARN están constituidas por "módulos", que pueden ser sustituidos unos por otros sin afectar el plegamiento global [164]. Por ejemplo, una hélice puede ser sustituida por otra de diferente secuencia, siempre y cuando tengan aproximadamente el mismo tamaño. Asimismo, diferentes tipos de interacciones terciarias entre regiones distantes pueden ser sustituídas unas por otras, si sus características estéricas y de conectividad son compatibles [165]. Claramente, estos módulos intercambiables corresponden a estructuras que se forman en el nivel más bajo de la jerarquía. Dado que módulos alternativos de igual tamaño son compatibles con un mismo plegamiento, el número de plegamientos o clases estructurales de un ARN formado por N nucléotidos es muy inferior al número de secuencias posibles (N^4) [166]. Por lo tanto, cada clase estructural está formada por numerosas secuencias alternativas, las cuales constituyen el "ensamble" [152]. La modularidad de ensambles es una herramienta valiosa para comprender la evolución de los ARN y por lo tanto es invocada en muchos estudios filogenéticos [86].

II.2 Ejemplos de estudios computacionales sobre el ARN

La literatura sobre simulaciones de DM aplicada al estudio del ARN cubre múltiples aspectos que incluyen: flexibilidad intrínseca, impacto de mutaciones específicas, catálisis enzimática, formación de complejos de ARN/proteína, modos de acción de antibióticos, predicción de estructuras y mecanismos de inhibición, así como el rol de las moléculas de agua y los diferentes iones sobre su estructura y dinámica [91]. La primera simulación de una molécula de ARN se reportó en 1984 [167]. Desde ese hito, los parámetros utilizados en las mismas fueron continuamente

reexaminados para mejorar su capacidad predictiva (Ver Sección III.3.1) [168, 169]. Con este fin se emplearon modelos de ARN pequeños [170–174]. Entre los modelos estudiados se incluyen tetra-*loops*, *stem-loops* y empalmes [175–182]. Conforme la técnica empezó a consolidarse, creció también el número de simulaciones enfocadas en modelos de mayor tamaño [183–187]. Entre los aspectos explorados en dichos trabajos se destacan la simulación del impacto de sustituciones puntuales para los procesos de plegado, la estabilidad y flexibilidad intrínseca del sistema, los procesos de hidratación, y la coordinación con cationes. Estos estudios fueron algunos de los primeros en establecer de manera teórica los tiempos involucrados en los eventos asociados a un nivel de jerarquización estructural en particular [152].

Otros elementos ribonucleicos ampliamente estudiados durante la primera década del siglo fueron segmentos de ARN virales como el elemento trans-activador de respuesta (TAR) o la Secuencia de Activación de la Dimerización ("DIS", por sus siglas en inglés) del virus del VIH [187–189]. Por su parte, existe un vasto número de trabajos abocados a la simulación de ribozimas. Particularmente, sus aspectos catalíticos, la influencia de iones, el agua, y de nucleótidos específicos, fueron algunos de los tópicos más frecuentemente explorados por DM [190–192]. Cabe destacar que aquellos trabajos en los que se exploran reacciones catalíticas no utilizan DM clásica, sino que requieren de un tratamiento cuántico electrónico, capaz de describir la ruptura y formación de enlaces (Ver Capítulo III).

La maquinaria de la traducción ribosomal fue otro de los sistemas predilectos en estudios basados en simulaciones de DM. Así, se implementaron diferentes modelos que incluían parte del ribosoma para comprender aspectos clave de los procesos de decodificación, acomodamiento, transferencia y translocación de péptidos [193–196]. También se obtuvieron importantes avances en el conocimiento sobre los mecanismos de unión de antibióticos [197–199]. Por último, cabe destacar que los constantes avances en las capacidades de cómputo ocurridos en este siglo, han permitido la realizar simulaciones en sistemas con millones de átomos. Entre estos estudios se encuentran los que modelaron el ribosoma completo, o moléculas enteras de ARNt [200]. Asimismo, se realizaron simulaciones que examinan las bases moleculares de la translocación y acomodamiento del ARNt [201–203].

II.2.1 Antecedentes sobre ARN pequeños (ARNp)

Los ARN pequeños (ARNp) tienen un rol fundamental en el control traduccional, particularmente en bacterias [204]. La implementación de técnicas de DM para el análisis de ARNp reguladores comenzó en la última década, por lo que la cantidad de ejemplos es relativamente escaso. Estas moléculas son de particular interés para esta tesis, ya que uno de los sistemas estudiados incluye al ARNp RsmZ (Ver Capítulos IV y V).

A modo ilustrativo, cabe mencionar que Hfq es una de las proteínas más ampliamente estudiadas por su influencia en el control traduccional de bacterias. Su función está, precisamente, regulada por moléculas de ARNp [205]. La Energía Libre de unión entre Hfq y uno de estos ARNp llamado OxyS se evaluó mediante simulaciones computacionales, encontrándose un muy buen acuerdo con los resultados experimentales previos [206]. En *Pseudomonas*, Hfq y la proteína de control de la represión catabólica (Crc) forman, junto con los ARNm blanco, un complejo de traducción-represión. La proteína Crc no se une a los ARN ni a Hfq por sí misma, sino a los complejos Hfq-ARN, fortaleciendo las interacciones de Hfq con los ARNm objetivo. Las bases moleculares para la interfaz ARN-Hfq y sus implicaciones en la unión de Crc fueron precisamente estudiadas por DM [207], lo que permitió identificar un nuevo bolsillo de unión de Hfq y una posible vía de plegamiento para el complejo ARN-Hfq-Crc. Estudios complementarios demostraron, además, que la concentración iónica en la célula es crucial para estabilizar las interacciones de apilamiento entre aminoácidos y nucleótidos de dicho complejo [208].

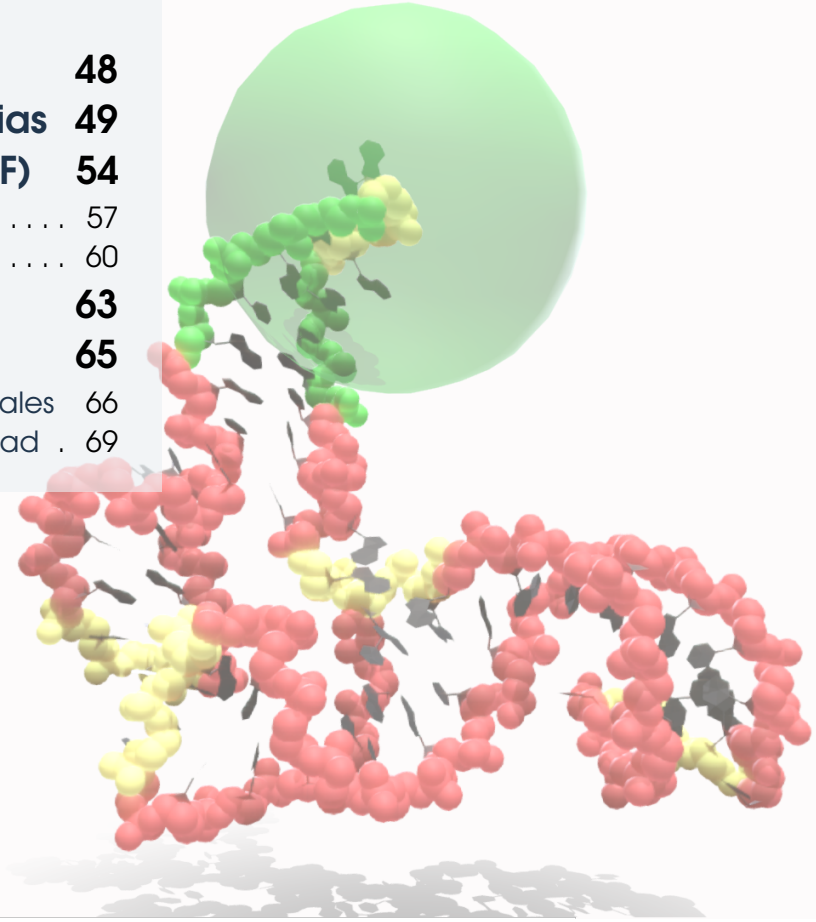
Finalmente, un estudio experimental reciente en plantas de arroz investigó las razones para su alto rendimiento y su eficiencia en el uso de Nitrógeno. Ese trabajo postuló que dicho fenómeno se puede explicar por la presencia de dos proteínas cuya traducción es regulada por dos ARNp de 24 nucleótidos [209]. La Referencia 209 reporta simulaciones de DM sobre estos ARNp que examinan la influencia que la temperatura tiene sobre sus estructuras, encontrando cambios en los *stem-loops* que permitirían su unión a los ARNm objetivo.

Las simulaciones presentadas en esta tesis fueron las primeras reportadas para

el estudio computacional del fragmento de 72 nucleótidos del ARNp RsmZ [2, 3], elucidadas por la Referencia 73.

Espacio para anotaciones

III.1 Fundamentos **48**
III.2 El modelo y las trayectorias **49**
III.3 Los campos de fuerza (CF) **54**
III.3.1 Campos de fuerza para ARN 57
III.3.2 Las etapas de una simulación 60
III.4 *Umbrella sampling* **63**
III.5 Análisis **65**
III.5.1 Análisis de Componentes Principales 66
III.5.2 Clusterización basada en densidad . 69



III. Conceptos básicos de Dinámica Molecular (DM)

Parte del desarrollo sobre los temas presentados en este Capítulo fueron incluidos en el libro "Bacterial Regulatory RNA: Methods and Protocols" de la serie "Methods in Molecular Biology" en el año 2023 [1].

Resumen de este capítulo En este capítulo se presentan los fundamentos del conjunto de métodos que se agrupan bajo el nombre de Dinámica Molecular (DM). En particular, se hace hincapié en las particularidades que los mismos adquieren cuando se aplican al estudio del ARN. El objetivo no es dar un desarrollo detallado en cuanto a la descripción de los algoritmos, los cuáles son variados y complejos, sino presentar un marco conceptual que permita evidenciar los alcances y limitaciones de esta metodología. Además, se describe a las diferentes etapas requeridas para llevar a cabo una simulación de DM y se discute la técnica de muestreo mejorado conocida bajo el nombre de *Umbrella Sampling*, de particular relevancia para este trabajo. El capítulo termina con una breve explicación de los métodos de análisis utilizados en esta tesis, implementados para obtener información relevante sobre los ARN a partir de la inmensa cantidad de datos colectados durante su simulación.

III.1 Fundamentos

Cada estado de equilibrio de un sistema macroscópico puede ser caracterizado indicando los valores de un pequeño conjunto de variables tales como (N,V,E), (N,V,T) o (N,P,T). De acuerdo a la termodinámica estadística, a cada uno de estos estados macroscópicos le corresponde un ensamble formado por un número infinito de copias del sistema [210]. Estas poseen los mismos valores de las variables macroscópicas, pero difieren en las coordenadas y momentos de sus partículas, que definen su estado "microscópico". Finalmente, se establece que el resultado de la medición de cualquier parámetro observable del sistema que pueda expresarse en función de las coordenadas y momentos de sus partículas, $O(\mathbf{x},\mathbf{p})$, es igual al valor promedio de ese observable evaluado en el ensamble, $\langle O \rangle$,

$$\langle O \rangle = \iint O(\mathbf{x},\mathbf{p})\rho(\mathbf{x},\mathbf{p})d\mathbf{x}d\mathbf{q}. \tag{III.1}$$

En esta expresión $\rho(\mathbf{x},\mathbf{p})$ mide la densidad de probabilidad de las coordenadas y momentos de las partículas $\mathbf{x}$ y $\mathbf{p}$ en el ensamble. Esta densidad es una propiedad que depende de las variables macroscópicas fijadas.

Bajo estas suposiciones, el problema que se presenta para calcular alguna propiedad del sistema es cómo evaluar la integral de la Ecuación III.1, debido al gran número de variables de integración. Para enfrentar este problema, se han propuesto diferentes alternativas [211]. La que es relevante para la presente discusión consiste en reemplazar el promedio en el ensamble por un promedio en el tiempo, evaluado a lo largo de una larga trayectoria del sistema por el espacio de sus variables microscópicas,

$$\bar{O}_T = \frac{1}{T}\int_0^T O(\mathbf{x}(t),\mathbf{p}(t))dt. \tag{III.2}$$

En principio, tal como lo establece la ecuación III.2, el valor del promedio temporal depende de la longitud de la trayectoria, T. Sin embargo, de acuerdo con la hipótesis ergódica [212], a medida que T tiende a infinito el promedio temporal se vuelve independiente de T y aproxima al promedio en el ensamble,

$$\lim_{T\to\infty} \bar{O}_T = \langle O \rangle. \tag{III.3}$$

Tal como su nombre lo establece, la hipótesis ergódica no tiene demostración. Es una hipótesis. No obstante, cabe destacar que no es sólo una suposición necesaria para poder realizar los cálculos, sino que también es razonable para los sistemas moleculares como los estudiados en esta tesis [213]. Estos sistemas sólo pueden acceder a un volumen acotado de su espacio de coordenadas y momentos, o espacio de las fases. Dado que están encerrados en ese volumen, es plausible asumir que tarde o temprano lo habrán recorrido en su totalidad [213], visitando el entorno de cada punto con una frecuencia acorde a su densidad de probabilidad, $\rho(\mathbf{x},\mathbf{p})$.

En la actualidad, se agrupa bajo el nombre de DM a un conjunto de técnicas de simulación computacional que tiene su fundamento (o su origen) en la hipótesis ergódica [214]. En un cálculo DM se busca determinar trayectorias de las moléculas bajo estudio mediante la implementación de las ecuaciones de movimiento pertinentes. Estas ecuaciones están basadas en las leyes de la mecánica newtoniana [91, 215]. No obstante, en la mayoría de los casos, las ecuaciones de Newton son modificadas para que el sistema tenga una temperatura constante en vez de una energía constante [211, 216]. En esta tesis se utiliza la llamada "DM clásica". En ella, se asigna una masa puntual a cada átomo o ión del sistema mientras que las interacciones entre los mismos se evalúan mediante una función analítica de sus posiciones [217]. Esta función recibe el nombre de CF, y se discutirá en la Sección III.3 de este Capítulo.

III.2 El modelo y las trayectorias

Para realizar una simulación DM primero es necesario definir el modelo a considerar, así como a las ecuaciones que se emplearán en la propagación de sus coordenadas iniciales. En simulaciones de biomoléculas, el modelo típicamente consiste de un cubo o un octaedro en cuyo centro se ubica la o las biomoléculas que se quieren simular [217]. En el caso que la molécula posea una carga neta, se agregan iones previo a la solvatación del sistema con el fin de neutralizar dicha

carga, con posiciones acordes al potencial electrostático de la macromolécula. Para el agregado del resto de los iones, que resultan necesarios para reproducir las concentraciones fisiológicas del medio a simular, se reemplazan estocásticamente moléculas aleatorias de solvente luego de llevar a cabo la solvatación del sistema.

La dependencia de la simulación con la estructura inicial es particularmente relevante en el caso de las moléculas de ARN [218]. Debido a que el espacio configuracional de las biomoléculas es de gran tamaño, si se comienza una simulación desde una configuración inapropiada, la equilibración del sistema puede demandar un tiempo enormemente largo. Caso contrario, puede que no lleguen a observarse nunca las configuraciones relevantes para el aspecto que se quiere analizar o que, en el conjunto de datos colectados, tengan un peso mucho menor que el que deberían tener. Lo ideal es contar con una estructura inicial de la biomolécula que haya sido determinada experimentalmente, tales como las que se pueden descargar del *Protein Data Bank* (PDB) (https://www.rcsb.org/).

Las estructuras de proteínas que se pliegan de formas definidas pueden ser predichas mediante el algoritmo Alpha Fold [219–221]. Este algoritmo resulta eficiente dado que el 97,89 % de las 208.066 entradas del PDB, las cuáles son empleadas como *set* de entrenamiento para la predicción, corresponden a proteínas o complejos formados por ellas [222]. Para el caso particular de las moléculas de ARN, la capacidad predictiva de este tipo de algoritmos está limitada por el hecho de que las estructuras determinadas experimentalmente representan solo una pequeña porción de la información disponible, pese a que el número de moléculas de ARN humanos supera al de proteínas en más de un orden de magnitud [223, 224]. Dada su baja representación en las bases de datos estructurales más generales, existen repositorios de estructuras de ácidos nucleicos (http://ndbserver.rutgers.edu/) y específicamente de ARN (http://rna.bgsu.edu/rna3dhub/pdb y https://rnasolo.cs.put.poznan.pl/).

El costo computacional de una simulación de DM depende del número de átomos del sistema considerado. Por lo tanto, es claro que las dimensiones del cubo u octaedro a utilizar son típicamente pequeñas [216]. Esto tiene como consecuencia que la relación entre la superficie del sistema y su volumen sea mucho mayor

que la de los sistemas macroscópicos que se busca emular. Debido a que las moléculas en la interfase se comportan de manera distinta que las ubicadas en el seno de la solución, el pequeño tamaño de la caja tiene como consecuencia que el sistema muestre un comportamiento artificial [225]. Para evitar este problema, las simulaciones de DM utilizan las llamadas condiciones periódicas de contorno. En las mismas, el sistema simulado está rodeado en todas las direcciones por sistemas idénticos [226]. De esta manera, los átomos y moléculas de la caja central (que es la que efectivamente se simula) interactúan con sus vecinos más cercanos, ya sea de la misma caja o de las cajas adyacentes. De esta manera se evitan los efectos de borde sin aumentar demasiado el costo computacional de la simulación.

Como se mencionó anteriormente, en DM clásica las ecuaciones de propagación se basan en las ecuaciones de Newton. Cuando las posiciones de las partículas se indican a partir de sus respectivas coordenadas cartesianas, estas ecuaciones adquieren la siguiente forma,

$$\frac{\partial \alpha_i}{dt} = v_{i,\alpha} \tag{III.4}$$

$$\frac{\partial v_{i,\alpha}}{dt} = -\frac{1}{m_i}\frac{\partial V}{\partial \alpha_i}, \tag{III.5}$$

donde α_i, con $\alpha = x$, y o z, denota alguna de las coordenadas de la partícula i, $v_{i,\alpha}$ es su velocidad en la dirección correspondiente, m_i es su masa y V es la energía potencial del sistema que se calcula con el CF (ver Sección III.3). Es importante destacar que estas ecuaciones son válidas cuando V no depende de las velocidades, como ocurre con todas las simulaciones descritas en esta tesis y en la mayoría de los cálculos DM de biomoléculas [227].

Si las ecuaciones III.4 y III.5 se propagan numéricamente utilizando un paso de integración apropiado, la energía del sistema se mantiene constante [216]. Sin embargo, en la mayoría de las simulaciones de DM, se busca emular situaciones en las que el sistema intercambia calor con un foco térmico a fin de mantener su temperatura constante. Para ello, es necesario modificar las ecuaciones III.4 y III.5 mediante la utilización de algoritmos que reciben el nombre genérico de

"termostatos" [216]. El número de termostatos disponibles es muy grande y cada uno de ellos tiene ventajas y desventajas que le son propias. En las simulaciones reportadas en esta tesis, se utilizó el termostato de Langevin [228], en el cual se altera la ecuación III.5 de cada partícula mediante,

$$\frac{\partial v_{i,\alpha}}{dt} = -\frac{1}{m_i}\frac{\partial V}{\partial \alpha_i} - \gamma v_{i,\alpha} + \sigma\xi, \tag{III.6}$$

donde el término $-\gamma v_{i,\alpha}$ da cuenta de una fuerza viscosa que siempre se opone al movimiento de la partícula considerada acorde a su velocidad. En particular, γ representa el coeficiente de fricción de esa fuerza, y determina cuán rápido se equilibran las partículas con el baño térmico. Asimismo, el término $\sigma\xi$ representa una fuerza aleatoria. Por un lado, ξ introduce esa aleatoriedad y se caracteriza por su valor medio cero y una varianza proporcional a la temperatura. De acuerdo con el teorema de fluctuación-disipación [229], las intensidades de estas fuerzas están relacionadas por σ, tal que,

$$\sigma^2 = 2\gamma m_i k_B T, \tag{III.7}$$

por lo tanto, al propagar el movimiento de una partícula i con una ecuación como III.6, la partícula está interactuando con un baño cuya temperatura es T. Este intercambio hace que dicha partícula también adquiera esa temperatura al alcanzar el equilibrio [230]. Así, estos términos simulan la interacción de la partícula considerada con un baño de partículas que, en lugar de ser tratadas explícitamente, son incluidas de acuerdo a sus efectos macroscópicos [216, 228].

El algoritmo de Langevin tiene dos ventajas importantes. Una de ellas es que no aumenta el costo computacional de manera apreciable [231]. La segunda y más importante es que efectivamente hace que el sistema reproduzca la distribución de velocidades de la T seleccionada [228]. En otras palabras, no sólo la energía cinética de las partículas se corresponde con la T requerida (esto es, que la suma de sus energías cinéticas iguala $3/2Nk_BT$), sino que la distribución de velocidades sigue una ley de Maxwell-Boltzmann [216]. Este aspecto es esencial para el estudio de procesos activados, es decir, de procesos que para ocurrir requieren superar

una barrera de energía. La principal desventaja de este algoritmo, es que no es determinista porque emplea fuerzas aleatorias. Debido a los cambios aleatorios en las velocidades de las partículas, el termostato de Langevin no es apropiado para el estudio de propiedades dinámicas de la biomolécula, tales como el coeficiente de difusión [216], pero no tiene mayores consecuencias para los análisis realizados en este trabajo de tesis.

Además de querer simular situaciones experimentales de T constante, muchas veces se quiere emular una presión constante. Para ello se hace uso de algoritmos que reciben el nombre de "barostatos". Ellos modifican las dimensiones de la caja simulada a fin de que se alcance la presión (y por ende, la densidad) buscada. De no utilizarse un barostato, el sistema muestrea un ensamble de (N,V,T) constantes (ensamble canónico), mientras que si se lo aplica, muestrea un ensamble con (N,P,T) constantes (ensamble isotérmico-isobárico) [232]. Para simulaciones de fases condensadas, como es el caso de las biomoléculas en solución, la diferencia entre V o P constantes no son significativas, siempre y cuando la densidad del sistema sea la misma en ambos casos. En cambio sí es muy influyente el uso de un termostato, sin el cual el sistema muestrearía un ensamble microcanónico (N,E,T) [232].

Finalmente, otra de las modificaciones que se introducen a las ecuaciones de movimiento y que fueron utilizadas en esta tesis, consiste en la aplicación multiplicadores de Lagrange para fijar la longitud de los enlaces que involucran átomos de hidrógeno (ej. enlaces C-H, O-H, S-H) [216]. En cada paso de la integración, el método evalúa qué fuerza debe aplicarse sobre los átomos que forman el enlace para evitar que se modifique la distancia entre los mismos. El objetivo de esta estrategia es eliminar de la simulación las vibraciones del sistema con mayor frecuencia [233]. De esta manera, es posible aumentar el paso de integración hasta 2 fs sin que se introduzca un error significativo a la trayectoria [234]. Al igual que con los termostatos y barostatos, existen diferentes algoritmos que aplican estas restricciones. En las simulaciones realizadas para este trabajo de tesis se utilizó el algoritmo denominado "SHAKE" [235], el cual evalúa el efecto de las fuerzas virtuales de Lagrange mediante un procedimiento iterativo.

III.3 Los campos de fuerza (CF)

Para simular la evolución temporal de un sistema molecular, es necesario propagar la ecuación de Schrödringer dependiente del tiempo utilizando un operador hamiltoniano que tenga en cuenta todos los núcleos y electrones del sistema, los cuales interaccionan entre sí por fuerzas de Coulomb [236]. Este tratamiento es absolutamente general, en el sentido de que es válido tanto para una molécula de H_2 como para el ARN o la proteína más grande que se quiera estudiar [237]. Además, está libre de aproximaciones, salvo por efectos relativistas que son despreciables para átomos del primer y segundo período de la tabla periódica [238]. El problema con este tratamiento es que es inabordable excepto para sistemas muy pequeños, compuestos por pocos átomos de pocos electrones [217].

Para superar la dificultad mencionada en el párrafo anterior, las simulaciones de DM utilizan una serie de simplificaciones. La primera de ellas es invocar la aproximación de Born-Oppenheimer para separar el movimiento electrónico del movimiento nuclear [239]. Bajo esta aproximación, los núcleos del sistema se mueven sobre una Superficie de Energía Potencial (SEP) determinada por los electrones. La segunda aproximación consiste en simular el movimiento de los núcleos sobre esta SEP mediante ecuaciones clásicas, como se discutió en la Sección III.2 [240]. Finalmente, la tercera gran aproximación consiste en calcular la SEP sobre la que se mueven los núcleos mediante una función empírica y relativamente sencilla, llamada CF [215, 241, 242].

La variedad de CF disponibles para simulaciones de DM es vasta [243]. Los mismos tienen dos características que deben ser claramente diferenciadas: la forma funcional del campo y sus parámetros [216]. Con una misma forma funcional, diversos campos son posibles si se utilizan parámetros diferentes. En las simulaciones reportadas en esta tesis, se emplearon los CF del paquete AMBER [233]. Las especificaciones de los mismos se darán en los capítulos pertinentes, y su forma funcional se muestra en la Figura III.1.

A cada enlace covalente del sistema, este CF le asigna un potencial armónico con constante de fuerza k_{r_i} y posición de equilibrio $r_{i,eq}$. También se asigna un potencial armónico a cada ángulo entre enlaces covalentes adyacentes. En este

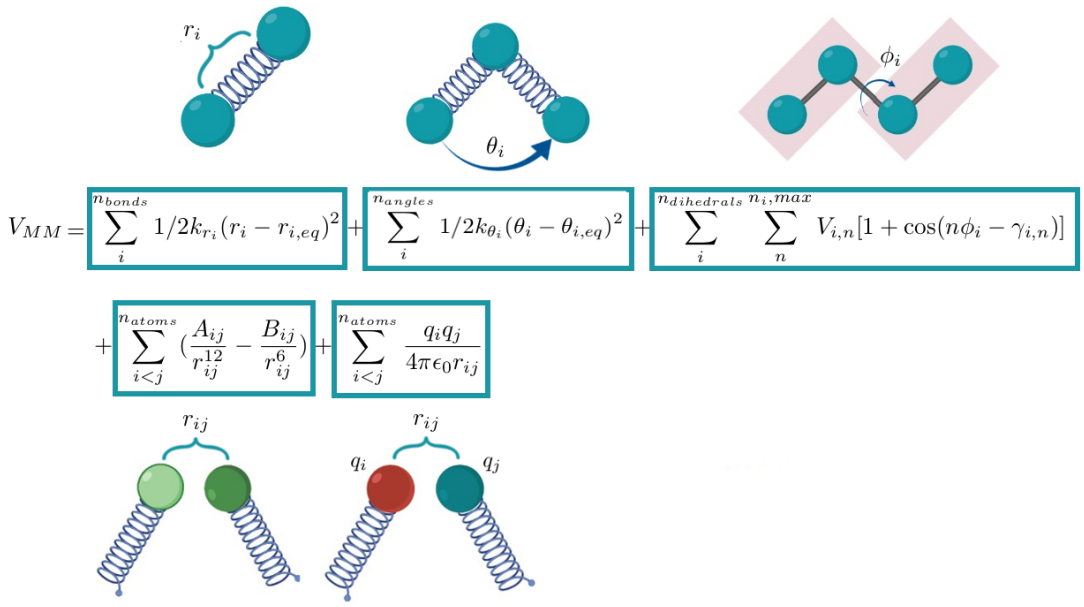


Figura III.1: Forma funcional de los CF utilizados en este trabajo de tesis. Imagen tomada de la Referencia 1.

caso las constantes de fuerza y posiciones de equilibrio están dados por k_{θ_i} y $\theta_{i,eq}$, respectivamente. Las rotaciones de ángulos diedros son tratadas mediante expansiones de Fourier cuyos coeficientes son los parámetros $V_{i,n}$. Idealmente, se requieren varios términos para una descripción precisa de las barreras encontradas en estas rotaciones. Sin embargo, por simplicidad, normalmente se emplea un solo término.

Las interacciones descritas hasta aquí reciben el nombre de interacciones ligadas, mientras que las que aparecen en las últimas dos sumatorias de la Figura III.1 se denominan interacciones no-ligadas [216]. Entre las no-ligadas se encuentra, en primer lugar, un potencial de Lennard-Jones que da cuenta de las atracciones por fuerzas dispersivas así como de las repulsiones de Pauli. Para cada par de átomos i y j, los parámetros correspondientes, $A_{i,j}$ y $B_{i,j}$, se calculan a partir de parámetros atómicos siguiendo las reglas de combinación de Lorentz-Berthelot [244, 245]. La última sumatoria describe las interacciones electrostáticas en el sistema. Para ello, se asigna a cada átomo del mismo una carga parcial q_i, la cual interacciona con el

resto de las cargas parciales mediante una fuerza de Coulomb [246].

Es necesario enfatizar que, si bien todos los términos que aparecen en el CF parecen intuitivamente correctos, muchos de ellos no corresponden a ninguna magnitud físicamente observable ni pueden ser determinados teóricamente sin ambigüedad [247]. Entre los parámetros que sí se corresponden con magnitudes observables se encuentran los correspondientes a los enlaces y ángulos entre éstos, que pueden determinarse a partir de información estructural obtenida por cristalografía de rayos X, así como de espectroscopía infrarroja y Raman [248]. Asimismo, los parámetros de Lennard-Jones se ajustan para que las simulaciones de DM tengan la capacidad de predecir las densidades y calores de vaporización de diversas sustancias [249]. Sin embargo, las cargas parciales y las barreras para las torsiones no son magnitudes directamente observables y, por ende, son mucho más difíciles de determinar [250]. Más aún, estas interacciones tienen una gran influencia sobre el comportamiento de las biomoléculas en solución.

En los CF de AMBER, las cargas parciales de los núcleos se determinan a partir de cálculos de estructura electrónica [233]. En estos cálculos, primero se determina el potencial electrostático sobre una grilla de puntos alrededor de la molécula considerada, sumando la contribución puntual de los núcleos más el efecto de la nube electrónica a su alrededor. Finalmente, se asignan cargas parciales fijas sobre cada uno de los átomos de la molécula de manera tal que reproduzcan, con el menor error posible, los potenciales de los puntos de la grilla [251, 252]. Esta estrategia tiene la virtud de reproducir las interacciones electrostáticas entre grupos moleculares de manera relativamente fidedigna para el bajo costo computacional que implica [216].

No se debe perder de vista que cuando una molécula interacciona con otra, las nubes electrónicas de ambas se distorsionan. Por lo tanto, el potencial eléctrico que ellas generan a su alrededor cuando están aisladas es diferente del que producen cuando están interactuando entre sí. Cuanto más deformable sea la molécula (es decir, cuanto mayor sea la polarizabilidad de los átomos que la conforman), más importante es este efecto y peor es la descripción que se obtiene mediante la simplificación del uso de cargas parciales [253]. Desde hace ya algunos años,

existen CF que tienen en cuenta los efectos de la polarizabilidad molecular [86]. Sin embargo, la utilización de los mismos aumenta enormemente el costo computacional del cálculo. Esto tiene como consecuencia que no se puedan alcanzar tiempos de muestreo suficientemente largos, excepto para moléculas pequeñas. En otras palabras, la exactitud que se gana en una mejor representación de las interacciones intermoleculares se pierde por un pobre muestreo [253].

Los últimos parámetros de un CF en ser determinados son los correspondientes a las torsiones [250]. Para su obtención se pueden seguir dos estrategias. La primera y más obvia consiste en utilizar cálculos cuánticos de manera de evaluar la energía electrónica para distintos valores del ángulo diedro en cuestión [254]. Para los mismos ángulos, se calcula la energía del CF sin la contribución de diedros, lo que hace que la diferencia entre ambas curvas sea significativa. Finalmente, se determina el o los valores de la expansión de Fourier del diedro considerado de manera tal de minimizar la diferencia entre ambas curvas [254]. Esta estrategia permite obtener una primera aproximación al valor de los parámetros de los ángulos diedros y funciona adecuadamente para moléculas pequeñas. Sin embargo, es deficiente para la determinación de los diedros del esqueleto de proteínas o de los ácidos nucleicos. Por lo tanto, para estos ángulos, los parámetros torsionales se han actualizado continuamente, de manera tal que las simulaciones de DM tengan la capacidad de reproducir las estructuras de grandes conjuntos de proteínas o ácidos nucleicos con el menor error posible [216]. A medida que las técnicas experimentales se desarrollaron incorporando más estructuras o estructuras determinadas en diferentes condiciones, los parámetros del CF para los diedros debieron ser adaptados. De hecho, si se analiza la evolución de los CF, se encuentra que de una versión a otra, las principales modificaciones están en este conjunto de parámetros [171, 250].

III.3.1 Campos de fuerza para ARN

Los CF de los ARN tienen un menor desarrollo, y por ende están sujetos a mayores errores y artefactos que los CF de proteínas (al menos los de proteínas plegables). Hay dos factores fundamentales que contribuyen a este estado de situación. El primer gran problema que enfrentan los CF de ARN es el efecto de no considerar

la polarizabilidad de los átomos [243]. Esto se debe a que el esqueleto de los ARN está compuesto por átomos de Fósforo, los cuales son altamente polarizables. Por otra parte, la no polarizabilidad en los CF utilizados en esta tesis incide sobre el modelado de correlaciones electrón-electrón en el apilamiento π–π de bases aromáticas [255]. Esta dificultad es incluso más marcada si la simulación contiene iones Mg^{+2}, debido a que estos son altamente polarizantes [243]. Otro problema asociado a la incorporación de iones Mg^{+2} es que sus concentraciones son muy bajas si se busca reproducir las condiciones fisiológicas [86]. En una simulación DM, esta baja concentración se traduce en un número muy pequeño de iones en la caja de simulación, lo que dificulta la equilibración de los mismos. En general, se sostiene la idea de que es mejor obviar la presencia de estos iones, excepto que se sepa que alguno de ellos cumple un rol estructural importante [243]. En tal caso, el ion Mg^{+2} se ubicará en el sitio donde está coordinado, donde típicamente permanece durante toda la simulación, sin equilibrarse con el resto de los iones de la solución.

El segundo problema proviene de las dificultades para determinar las estructuras de los ARN, las cuales a su vez se originan en la mayor diversidad conformacional de estas moléculas [218]. Así, la falta de información experimental suficiente, sumada a las dificultades para describir situaciones experimentales más variables y diversas, ha tenido como consecuencia que la capacidad predictiva de los CF de ARN sea menor que los de las proteínas plegables [91]. En años recientes, ha habido importantes avances al respecto y diversos grupos de investigación están enfocados en abordar estas dificultades, como se describe en los párrafos siguientes.

El CF de AMBER ff94 introdujo las torsiones específicas del ARN [256]. Estas se estimaron ajustando un perfil de energía QM para el fosfato de dimetilo, con el objetivo de usarse como referencia para el esqueleto polinucleotídico [243]. Lo propio ocurrió con el grupo ribosa, para garantizar una baja barrera de transición entre las conformaciones C2' endo y C3' correspondientes a las hélices de forma A y B. Sin embargo, las primeras simulaciones de DM aplicadas sobre ARN demostraron limitaciones importantes de los CF hasta entonces disponibles [171, 172]. Esto desencadenó la búsqueda de mejoras y constantes refinamientos en los

últimos años, los cuales continúan al día de hoy en desarrollo [91].

Las mejoras más recientes en las implementaciones de AMBER para la simulación de ácidos nucleicos consisten precisamente en el refinamiento de las torsiones α/γ en el esqueleto de estas moléculas [257], así como el posterior reajuste de las torsiones ε/ζ incluido en el CF BSC1 [258]. Con respecto al modelado de ARN, la primera mejora significativa se produjo en 2010, cuando se realizó un escrutinio sistemático de la torsión glucosídica χ para cada uno de los cuatro nucleósidos de ARN en el CF ff99-χYIL [259]. Al emplear un enfoque similar, todas las torsiones de ARN se reajustaron para una mejor reproducción de las estructuras obtenidas por RMN para el tetrámero r(CCCC) [260].

La posterior revisión del parámetro asociado al ángulo χ permitió eliminar estructuras espurias para el ARN en su forma "A" [261]. En dicha revisión se encontró que descripciones defectuosas de éste ángulo tienen como consecuencia una completa desestabilización de la estructura nativa de los ARN de secuencia larga. El CF que incluye esta última parametrización es ahora el modelo oficial de AMBER para ARN, y se conoce como ff99 bsc0$_{\chi OL3}$ [233]. Este fue, precisamente, el CF utilizado en esta tesis para la simulación de RsmZ. Al día de hoy, continúa el constante desarrollo de CF abocados a una mejor descripción de los ácidos nucleicos. Algunos de ellos parten de las últimas actualizaciones de AMBER para su refinamiento [262–267], mientras otras se desarrollan independientemente de este paquete de simulación en particular [268].

A pesar de que AMBER no exige el uso de ningún modelo de agua específico, se ha demostrado recientemente que el comportamiento de las simulaciones de ARN depende en gran medida del modelo de agua utilizado [269, 270]. En particular, la Energía Libre de apareamiento entre bases está intrínsecamente ligada a la fuerza de la interacción que éstas a su vez establecen con el agua [243]. Para este trabajo de tesis se empleó el modelo TIP3P, en el cual el agua es representada por tres cargas puntuales fijas correspondientes a las cargas del Oxígeno y los dos Hidrógenos [271]. Para el Oxígeno se incluyen además parámetros de Van der Waals. Los parámetros de este modelo se desarrollaron para reproducir propiedades macroscópicas del agua líquida tales como la densidad, el momento dipolar, la función de distribución

radial, el coeficiente de difusión, y su calor de vaporización. Existen otros modelos de agua, más sofisticados que TIP3P. Uno de ellos, que ha acaparado la atención en años recientes, es el modelo OPC, que reproduce mejor que TIP3P las propiedades electrostáticas del agua [272, 273]. Sin embargo, OPC es un modelo de cuatro puntos. Por este motivo, su implementación tiene un costo computacional más elevado. Así, al tener que decidir sobre cómo implementar una simulación de DM, nuevamente se presenta la disyuntiva de elegir entre el CF más exacto o el más eficiente (que permite realizar un mejor muestreo).

Habiendo mencionado las dificultades encontradas en el desarrollo de los CF para simular ARN, cabe destacar que los utilizados en este trabajo de tesis han demostrado su capacidad para reproducir las características estructurales de un conjunto significativo de ARN (Ver Sección II.2.1). Seguramente, estos campos pueden ser mejorados o refinados. Pero esto no invalida su capacidad para clarificar los aspectos más destacados y típicos de esta clase de moléculas.

III.3.2 Las etapas de una simulación

Aún cuando la estructura inicial para el modelo a simular parta de un experimento, es posible que la misma tenga desperfectos relacionados con la metodología utilizada para su elucidación [274]. Por otra parte, es habitual que en estas estructuras estén ausentes algunos átomos o residuos, los cuáles son agregados *in silico* mediante el uso de plantillas en las etapas de preparación del modelo [275]. En dicha instancia, además, es que se construye la caja de solvatación y se produce el agregado de iones. En cualquiera de los casos, esto puede producir la superposición entre pares de átomos o estructuras muy tensionadas que es necesario minimizar para evitar que el sistema colapse al iniciar la simulación.

La etapa de minimización es anterior a la propagación de las coordenadas iniciales del sistema y permite alcanzar uno de los mínimos locales posibles de la molécula en el CF elegido [241]. Para esto, el primer paso posterior a la generación del modelo es aplicar un algoritmo de minimización de energía en el que el sistema se mueve a lo largo de la dirección definida por el gradiente de la energía potencial. En el caso de AMBER, este algoritmo suele dividirse en dos etapas que utilizan

métodos distintos: de descenso más pronunciado, y de gradiente conjugado. Para el detalle sobre ambos métodos se recomienda la lectura de la Referencia 233.

En este punto, el modelo minimizado ya posee coordenadas iniciales aceptables, se encuentra embebido en una solución que contiene iones y tiene un volumen definido. Sin embargo, como los átomos no tienen velocidades asignadas, la situación es equivalente a poseer una temperatura igual a 0K en términos clásicos. En este punto, es necesario asignar esas velocidades acorde a una distribución dada, y hacer uso de los ya mencionados termostatos (ver Sección III.2) para garantizar que la misma se mantenga a pesar de que las partículas cambien constantemente sus velocidades [276]. Este cambio de temperatura debe realizarse de manera gradual para evitar que una alteración abrupta en las velocidades del sistema desestabilice la estructura del modelo [233].

El último ajuste antes de la etapa de producción de la DM es la equilibración del sistema. Esto permite que el mismo se asemeje a las condiciones experimentales al relajar su volumen hasta que su presión sea la deseada (típicamente $1.0atm$), la cual se mantendrá constante durante el resto de la simulación. Para esto, se hace uso de los también ya mencionados barostatos, que escalan el volumen del sistema para alcanzar la presión objetivo [277] (Ver Sección III.2). Este último ajuste también cambia la densidad del sistema, que debería fluctuar alrededor de 1,0 g/mL para simulaciones en soluciones acuosas.

Las variables configuradas para ejecutar la etapa de equilibración no suelen diferir significativamente de las que se emplean en la posterior etapa de producción. En muchos casos, incluso, esta etapa está contemplada dentro de la misma simulación, y se determina el punto a partir del cual se considera que el sistema está equilibrado, empleando el resto de la trayectoria para el análisis de los resultados. Uno de los requisitos indispensables para considerar que el sistema se encuentra equilibrado es que la temperatura, la densidad y la energía del mismo oscilen en torno a valores constantes. A esto se debe agregar la constancia de los valores de *Root Mean Square Deviation* (RMSD) respecto de la estructura inicial del sistema [233]. Este parámetro mide la distancia promedio entre los átomos correspondientes a dos estructuras alternativas, después de que las mismas han sido alineadas. Sin

embargo, es importante distinguir entre estas dos etapas dado que el cambio en el volumen del sistema puede producir perturbaciones momentáneas que no necesariamente están relacionadas a la función biológica de la molécula estudiada, por lo que esta porción de la trayectoria no debería utilizarse para explicar su comportamiento en la naturaleza.

La etapa de producción consiste en la sucesiva integración de la ecuación mostrada en la Figura III.1 con las condiciones generales del sistema ya alcanzadas, utilizando para cada paso las coordenadas del sistema en ese momento. De esta forma, los movimientos del sistema son descritos como series de *snapshots* con resoluciones del orden de los femtosegundos. Una vez obtenidas las nuevas posiciones y velocidades, se actualizan los valores de la energía potencial y las fuerzas (y, por lo tanto, de las aceleraciones) para propagar nuevamente los valores de las posiciones y las velocidades [233].

Una vez finalizada la producción, se obtiene una trayectoria de conformaciones que, analizadas en conjunto, permiten inferir aspectos dinámicos y estructurales relevantes para el rol biológico de la molécula estudiada. No obstante, existen casos en los cuáles una única trayectoria no permite describir una porción representativa del espacio conformacional de una molécula, sino a una porción reducida en torno a un mínimo. Por este motivo, se acostumbra la ejecución de varias réplicas de la simulación, en cada una de las cuáles se parte de velocidades iniciales asignadas de forma aleatoria sobre cada átomo que en conjunto describen una misma distribución respecto a la temperatura deseada. Si el número de réplicas y los tiempos de simulación son suficientes, el conjunto de trayectorias concatenadas permitiría describir una porción representativa del espacio conformacional del sistema [91]. Cuando se cumple este objetivo, se dice que la simulación es convergente.

Alcanzar la convergencia sobre un conjunto de simulaciones es una tarea compleja, especialmente cuando se trata del ARN dado su vasto espacio conformacional [278]. Por otra parte, vale recalcar que algunos cambios conformacionales asociados con este tipo de moléculas ocurren en tiempos que exceden a los alcances de las simulaciones de DM (ver Sección II.1.3). Por lo tanto, es importante circunscribir las conclusiones que emergen de los resultados de una simulación a los límites que

tiene la técnica, evitando inferir de manera taxativa lo que excede a sus alcances.

En este sentido, cabe mencionar que el desarrollo creciente en materia tecnológica tuvo también un impacto significativo en la capacidad de cómputo para el campo de las simulaciones computacionales. Particularmente, la adaptación del paquete AMBER a su ejecución en placas de video ("*GPU*", por sus siglas en inglés) en el año 2016 permitió una reducción sensible en los tiempos de cálculo, logrando así obtener trayectorias más largas, en sistemas de mayor tamaño y tiempos menores de simulación [279]. Esto a su vez impacta sobre la precisión de los resultados, por mejorar la muestra estadística del sistema estudiado.

Pese a estos avances, persisten procesos a nivel molecular cuyos tiempos son aún inabarcables por las técnicas tradicionales de DM, dado que implican barreras de Energía Libre pronunciadas. No obstante, existen adaptaciones a esta metodología que permiten explorar eventos infrecuentes, como se detalla en la siguiente sección.

III.4 *Umbrella sampling*

Como se describió en la Sección II.1.3, las biomoléculas en solución pasan la mayor parte del tiempo en configuraciones cercanas a los mínimos de los pozos de sus SEL. Lo propio ocurre con las simulaciones de DM que intentan emular el comportamiento de esas biomoléculas. Típicamente, se observa que estas simulaciones permanecen en el pozo de la SEL al que pertenece la estructura inicial desde la cual fueron iniciadas [91]. Esto plantea un problema cuando se intenta estudiar procesos que requieren atravesar barreras elevadas de Energía Libre, como ocurre cuando un sustrato sale del sitio activo de una enzima [280], o cuando la biomolécula ejecuta un cambio conformacional [281]. A $T \sim 300K$, si la diferencia de Energía Libre que hay que superar está por encima de las $10kcal/mol$, lo más probable es que el evento que se quiere estudiar nunca sea reproducido por una simulación de longitud estándar [1]. Y en caso de ser observado, su ocurrencia sería demasiado baja como para permitir un análisis estadístico certero. Para enfrentar situaciones como estas, se han diseñado un conjunto de técnicas que se conocen con el nombre de "muestreo mejorado" [282].

Dado que los problemas de muestreo se originan en que los sistemas quedan

atrapados en pozos de Energía Libre, es posible mitigarlos mediante dos estrategias alternativas: o bien se incrementa la energía cinética del sistema aumentando T [282], o se agregan nuevos términos a la función de la energía potencial de manera tal que se reduzca la barrera a superar [283]. Cualquiera sea la estrategia seguida, luego de haber conseguido el muestreo en las condiciones ficticias seleccionadas (mayores T o potencial alterado), es necesario ajustar las ponderaciones de las muestras obtenidas a fin de que provean información sobre el proceso en las condiciones reales.

En este trabajo de tesis se utilizó la técnica de muestreo mejorado que se conoce con el nombre de "*Umbrella Sampling*" [283]. El objetivo fue simular la salida de un dímero de RsmE de uno de los sitios de unión de RsmZ, así como del constructo $hcnA_{20-mer}$ (ver Sección I.3). En las simulaciones *Umbrella Sampling* primero se debe elegir una coordenada de reacción apropiada (χ) que describa adecuadamente el proceso a estudiar. Luego se realizan simulaciones en las que el potencial provisto por el CF se modifica por el agregado de un potencial de sesgo,

$$V_{\text{ses}} = \frac{1}{2}k_{\text{ses}}(\chi - \chi_0)^2, \tag{III.8}$$

donde el efecto de este potencial es que las configuraciones del sistema oscilan (o tienden a oscilar) en un entorno de χ_0. Por lo tanto, a fin de recorrer todo el rango relevante de la coordenada de reacción se realiza un serie de simulaciones, cada una de las cuales tiene un valor diferente de χ_0, de manera tal de cubrir totalmente dicho rango. Cada una de estas simulaciones se conoce como "ventanas", y permite observar el comportamiento del sistema en una pequeña porción de los valores posibles de χ.

Una vez que se completaron todas las simulaciones, el rango de completo de χ se subdivide en un cierto número de cajas. Típicamente, este número es al menos 3 o 4 veces el número de ventanas. A partir de las muestras obtenidas de las simulaciones sesgadas, es posible determinar la probabilidad $P^{(n)}(\chi_i)$, de encontrar al sistema en la caja i, con centro en χ_i, durante la simulación de la ventana n. El paso final del cómputo consiste en desesgar estas probabilidades a fin de obtener las probabilidades correspondientes al potencial original $P(\chi_i)$. A partir de las mismas,

se puede calcular el PEL mediante la aplicación de la Ecuación II.4.

Lamentablemente, no es posible obtener las probabilidades no sesgadas a partir de las sesgadas de manera analítica, sino que sólo se puede alcanzar una solución numérica. Cada solución numérica está sujeta a errores, tanto de convergencia como debidas a valores inapropiados de los parámetros de *Umbrella Sampling* aplicados. Para evitar este tipo de errores, en este trabajo de tesis se emplearon tres algoritmos alternativos para obtener las $P(\chi_i)$ a partir de las $P^{(n)}(\chi_i)$. Ellos son el Método de Análisis de Histograma Ponderado (WHAM, por sus siglas en inglés), el Método de Análisis de Histograma Dinámico (DHAM, por sus siglas en inglés) y el Índice de Aceptación de Bennett (BAR, por sus siglas en inglés). Los fundamentos de los mismos pueden encontrarse en las Referencias 284, 285 y 286, respectivamente. Dado que los tres métodos utilizan estrategias distintas y son afectados de manera diferente por la elección de los parámetros de *Umbrella Sampling*, la coincidencia entre los PEL predichos por cada uno provee un reaseguro de la precisión de dichos perfiles.

III.5 Análisis

Una vez obtenida una o varias trayectorias de DM para el sistema a estudiar, es necesario analizar los datos obtenidos a fin de traducirlos en información inteligible y relevante para el problema en cuestión. Este análisis puede abarcar desde la determinación de parámetros simples y directos como distancias o ángulos entre grupos de átomos, hasta cálculos complejos obtenidos a partir de un entramado teórico más elaborado. El cálculo de constantes de unión, cambios de entropía o perfiles de Energía Libre, son ejemplos que pertenecen a este segundo grupo. Es importante destacar que el nivel de complejidad del análisis no necesariamente está correlacionado con el impacto de las observaciones que surgen de ellos. Por citar un ejemplo, uno de los hitos de la biología computacional moderna fue el advenimiento de Alpha Fold para la predicción de estructuras proteicas a partir de secuencias [219, 220]. Uno de los análisis empleados para evaluar la calidad de las estructuras predichas es el valor de RMSD entre estas estructuras y las obtenidas por técnicas experimentales. El análisis de RMSD es uno de los ensayos más sencillos y más

asiduamente utilizados en un trabajo típico de DM. Más allá de su simpleza, sirvió para determinar de manera contundente el potencial predictivo de Alpha Fold.

Entre otras técnicas de análisis sencillas, cabe mencionar al RMSF (por sus siglas en inglés *Root Mean Square Fluctuations*) como medida del desplazamiento a lo largo de la trayectoria de un átomo en particular, o grupo de átomos, en relación con una estructura de referencia [287]. Otros ensayos habituales son la identificación de motivos estructurales y su seguimiento a lo largo de la trayectoria [74], así como la construcción de mapas de contactos entre residuos (Ver Sección V.3).

Sin embargo, no siempre es posible sacar conclusiones interesantes apelando únicamente a los métodos más básicos de análisis. Esto se debe, fundamentalmente, a que las dimensiones del problema que se debe analizar es muy vasta. Por este motivo, resultan especialmente útiles los métodos que permiten identificar un número reducido de variables relevantes para describir el funcionamiento del sistema, como el que se describe a continuación.

III.5.1 Análisis de Componentes Principales

El análisis de componentes principales (PCA, por sus siglas en inglés) es un procedimiento muy utilizado en la ciencia de datos, donde se aplica con diversos propósitos [288]. Tiene también un uso muy extendido en el análisis de simulaciones de DM de biomoléculas [289]. En este contexto, su principal objetivo es construir un espacio vectorial de pequeñas dimensiones (típicamente menos de 20 grados de libertad), que permite describir las deformaciones/fluctuaciones con mayor amplitud de la biomolécula. La idea subyacente es que estas fluctuaciones son las que permiten a la misma cumplir con su rol biológico [288].

La primera implementación del PCA para analizar simulaciones de DM surgió de la observación de que, en una simulación de la Crambina, las fluctuaciones de sus átomos estaban fuertemente correlacionadas [290]. Poco tiempo después, Berendsen y colaboradores popularizaron el procedimiento al afirmar que el mismo permite definir el Espacio Esencial (EE) de las proteínas [291]. Este es un pequeño subespacio que contiene (o debería contener) todos los movimientos requeridos para su funcionamiento. El principal punto en estos dos trabajos consistió en notar

que solo unos pocos autovectores de la matriz de covarianza daban cuenta de la mayor parte de las fluctuaciones observadas en la trayectoria. Por lo tanto, solo este pequeño conjunto sería requerido para describir su dinámica "esencial", separándola de lo que podría considerarse el "ruido" de la simulación.

El primer paso para realizar un PCA es calcular la matriz de covarianza de las variables observables a incluir en el análisis. Los elementos de la matriz de covarianza, **C**, se calculan a partir de los datos colectados en una simulación mediante,

$$C_{ij} = \frac{1}{N}\sum_{k=1}^{N}(x_i^{(k)} - \langle x_i \rangle)(x_j^{(k)} - \langle x_j \rangle). \tag{III.9}$$

En esta expresión, $\langle x_i \rangle$ y $\langle x_j \rangle$ son los promedios de las coordenadas x_i y x_j durante la trayectoria, respectivamente, mientras que k es un índice que identifica las muestras y N es el número total de muestras colectadas. Las coordenadas utilizadas pueden ser de diferentes tipos. Las más utilizadas son las coordenadas cartesianas de los átomos del esqueleto de la molécula [292], pero otras opciones como ángulos diedros o distancias de contactos también pueden aplicarse [293]. Por último, cabe mencionar que si van a usarse como variables las coordenadas cartesianas de los átomos, el cálculo de la matriz de covarianza debe estar precedido por la eliminación de los efectos de la traslación y la rotación sobre las mismas, de manera tal que solo queden las deformaciones internas [289].

El paso siguiente del análisis es la diagonalización de la matriz de covarianza. Esto implica obtener la matriz **R** que, mediante una transformación de similaridad, convierte a **C** en la matriz diagonal Λ,

$$\mathbf{R^T C R} = \Lambda. \tag{III.10}$$

En esta expresión $\mathbf{R}^T$ es la transpuesta de **R**. Las columnas de **R**, $\mathbf{r}_i$, son los autovectores de **C**. Estos vectores forman una base ortonormal que expande el mismo espacio que la base cartesiana original. Si se proyecta alguno de los desplazamientos originales $(\mathbf{x^{(k)}} - \langle \mathbf{x} \rangle)$ sobre $\mathbf{r}_i$, se obtiene el valor del Componente Principal i

en la muestra k, $PC_i^{(k)}$. Las fluctuaciones de estos Componentes Principales son los elementos diagonales de Λ.

En principio, la base formada por las coordenadas originales y la de los autovectores de $\mathbf{C}$ son equivalentes. Además, la suma de las fluctuaciones de las coordenadas originales es igual a la suma de las fluctuaciones de los Componentes Principales. Esto surge del hecho de que ambas bases están relacionadas por una transformación de similaridad. Sin embargo hay una diferencia importante entre ambas que tiene consecuencias prácticas: mientras que las fluctuaciones totales están repartidas de manera más o menos equitativa entre todas las coordenadas originales, las mismas están concentradas en unos pocos Componentes Principales. Por lo tanto, esos pocos Componentes Principales son los que se requieren para describir los movimientos más importantes de la biomolécula, de acuerdo a lo observado en la simulación [289].

El resultado de un PCA puede visualizarse de diversas maneras. Una de las más sencillas e intuitivas es la representación en forma de película de los movimientos colectivos descritos por los autovectores importantes. Cuando las coordenadas originales son cartesianas, el programa de visualización VMD permite generar estas animaciones mediante la extensión llamada NMWIZARD [294]. También se pueden hacer ilustraciones de estos vectores mostrando cómo desplazan a los átomos incluidos en el análisis. Ambos procedimientos fueron utilizados en esta tesis para representar las principales deformaciones sufridas por RsmE y RsmZ durante el proceso de unión entre sí (Ver Sección V.4).

Otro uso muy extendido de PCA es el de construir SEL [288]. Para ello, y una vez obtenido el conjunto de los $\{PC_i^{(k)}\}$ para al menos dos vectores principales, se divide el rango de cada uno de ellos en una grilla con “cajas” de igual tamaño. Finalmente, se estima la probabilidad de cada centro de la grilla dividiendo el número de muestras que caen en la misma por el número total de muestras. Este cociente es una estimación de la probabilidad de hallar al sistema en ese punto y por lo tanto, mediante la aplicación de la Ecuación II.4, se puede calcular la Energía Libre correspondiente.

Cabe destacar que para la determinación de las SEL, así como para la aplicación

de algoritmos de clusterización como los que se describen en la próxima sección, el uso de PCA de coordenadas cartesianas es limitado [288]. Esto se debe a que, como se mencionó anteriormente, la implementación de estas coordenadas requiere eliminar el efecto de la translación y rotación global de la molécula antes de calcular la matriz de covarianza. Sin embargo, mientras que la eliminación de la traslación puede realizarse de manera analítica y exacta, la eliminación de la rotación es solo aproximada y debe efectuarse numéricamente [288]. Dado que la rotación y la vibración son movimientos que están acoplados, si la molécula es relativamente rígida (como es el caso de muchas proteínas globulares) es posible adosar a la misma un sistema de ejes coordenados y luego considerar a la rotación de estos ejes como la rotación global de la molécula [288]. Sin embargo, si la molécula ejecuta grandes cambios conformacionales, como ocurre con la mayoría de los ARN, el procedimiento resulta inadecuado [295]. Para esos casos, se recomienda hacer PCA de coordenadas internas que, por su propia definición, son independientes de las rotaciones globales. En particular, se ha descrito que el PCA de ángulos diedros del esqueleto, llamado dPCA, es apto para caracterizar las SEL de los ARN [296]. Dicha metodología fue aplicada en esta tesis para el análisis de RsmZ (Ver Sección IV.3).

III.5.2 Clusterización basada en densidad

El número de estructuras obtenidas en las simulaciones de DM es típicamente grande. Por este motivo, muchos análisis comienzan por agrupar estructuras que se consideran similares o relacionadas, ya sea geométrica o dinámicamente, a fin de obtener un número más manejable de situaciones a escrutar. Los algoritmos utilizados con este fin se conocen con el nombre genérico de métodos de clusterización.

En este trabajo de tesis, se utilizó un método de clusterización desarrollado por Stock y sus colaboradores [293]. Su implementación es algo más compleja que la del popular *k-means* [297]. Sin embargo, su mayor costo de implementación se justifica ya que define los *clusters* con un criterio físico, no sólo geométrico. Así, mientras *k-means* ubica la frontera entre *clusters* adyacentes de manera tal que

sea equidistante de ambos centros, el algoritmo de Stock ubica (o intenta ubicar) este límite en el máximo de la barrera de Energía Libre que separa los *clusters*. La aplicación del método tiene cuatro etapas:

1. Primero se reduce la dimensionalidad del espacio de conformaciones utilizando PCA.
2. Luego se calculan las distancias entre todas las estructuras de la muestra en el espacio de dimensionalidad reducida.
3. Utilizando todo el conjunto de distancias se calculan las densidades locales, con las cuales se estiman las Energías Libres.
4. A continuación, partiendo de las Energías Libres locales, se construyen los *clusters* mediante un proceso que tiene en cuenta tanto las similitudes geométricas entre los mismos como los valores de Energía Libre.

Dado que el uso de PCA para reducir la dimensionalidad del espacio de configuraciones se discutió en la sección precedente, en lo que sigue se discuten los puntos 2 a 4. La distancia entre la estructura i y la j en el espacio de dimensionalidad reducida, d_{ij}, se calcula simplemente como una distancia Euclidiana. Luego, la densidad local en torno a cada estructura se estima mediante,

$$P_R(i) = \sum_{j=1}^{M} \theta(R - d_{ij}), \tag{III.11}$$

donde $\theta(x)$ es la función escalón de Heaviside, M es el número total de muestras y R es el radio de una hiperesfera que contiene a las estructuras que contribuyen a la densidad local en la estructura i. Finalmente, con las densidades locales $P_R(i)$ se estima la Energía Libre de cada estructura mediante,

$$G_R(i) = -k_B T ln \left(\frac{P_R(i)}{P_R^{max}} \right), \tag{III.12}$$

donde P_R^{max} es la densidad local más alta a la cual, en virtud de la Eq. III.12, se le asigna una Energía Libre de cero. Esto evidencia que las densidades locales calculadas con la Eq. III.11 dependen del valor elegido para R. Para determinar el valor óptimo de este parámetro se calcula el conjunto de densidades locales

$\{P_R(i)\}$ para diferentes R. Con cada uno de estos conjuntos se evalúa una densidad de probabilidad de densidades locales. Es decir, se evalúa cuántos *frames* hay en los diferentes valores entre 0 y 1, y se divide por el ancho del intervalo utilizado. El máximo valor de R para el cual estas densidades de probabilidad tienen forma exponencial es el valor que debe utilizarse.

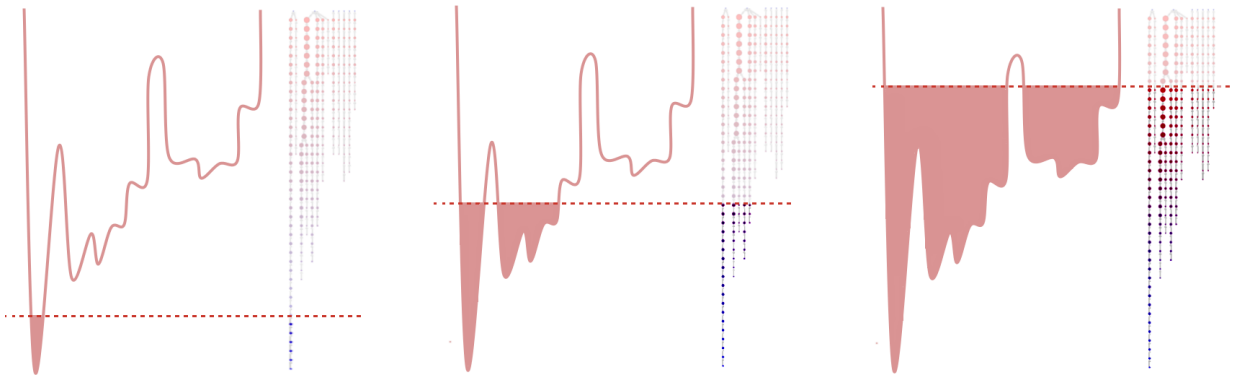


Figura III.2: Representación esquemática del algoritmo de clusterización basado en densidad propuesto por Stock y sus colaboradores [293]. Cada panel muestra la curva de Energía Libre total para el sistema, mientras la línea punteada señala el valor de corte para la incorporación de estructuras a cada *cluster*. A su vez, se muestra el cladograma que denota la estructura jerárquica de los *clusters* en forma de nodos. De esta manera, queda representado cómo el incremento gradual en la Energía Libre de corte tiene como consecuencia la incorporación de nuevas estructuras a cada *cluster*.

Habiendo estimado la Energía Libre local para cada *snapshot* de la trayectoria, los *clusters* se construyen mediante el siguiente procedimiento: En un primer paso se elige un valor relativamente bajo para la Energía Libre (típicamente $< 0.1k_BT$) y se seleccionan todas las estructuras con Energías Libres menores a este valor. El resto, por el momento, son descartadas. En un segundo paso, se agrupan en un mismo *cluster* todas las estructuras seleccionadas cuyas distancias sean menores que un valor de corte, d_{lump}. Esta asignación es transitiva de manera tal que si $d_{AB} < d_{\text{lump}}$, y $d_{BC} < d_{\text{lump}}$, entonces A, B y C se asignan a un mismo *cluster* aun cuando $d_{AC} > d_{\text{lump}}$. Así, un *cluster* se define de manera recursiva como el conjunto de estructuras con al menos un vecino cercano en el mismo conjunto. Por el contrario, dos conjuntos de estructuras son disjuntos y forman dos *clusters*

distintos si ninguna de las estructuras del primer conjunto está más cerca que el umbral de agrupamiento de alguna de las estructuras del segundo. El valor de d_{lump} sugerido en la Referencia 293 es $4\langle d_{\text{NN}}\rangle$, donde d_{NN} es la distancia entre vecinos más cercanos y los *brackets* denotan el valor promedio.

A continuación el valor de corte de la Energía Libre se incrementa de manera gradual. Con cada incremento, nuevas estructuras se incorporan al análisis y se agrupan en los *clusters* preexistentes de acuerdo al procedimiento descrito anteriormente. Al contener progresivamente más estructuras, los *clusters* crecen tanto en población como en el volumen que ocupan del espacio de dimensionalidad reducida. Eventualmente, ocurre que *clusters* originalmente diferentes se unen cuando la Energía Libre de corte supera la barrera entre los mismos. La Figura III.2 muestra una representación esquemática de este procedimiento.

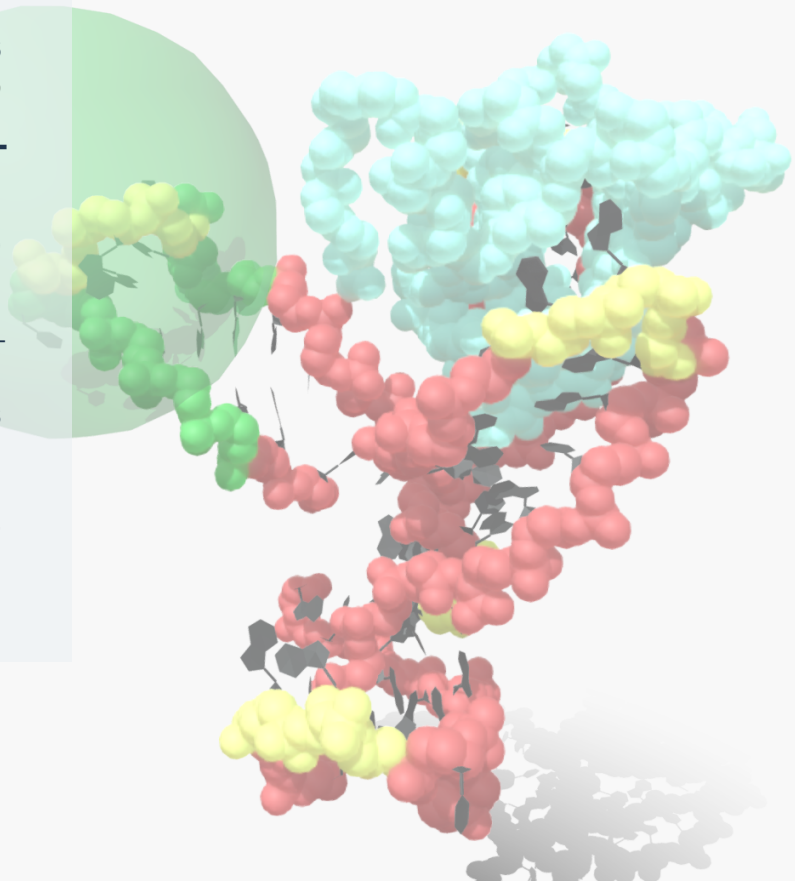
Este procedimiento genera una estructura jerárquica de nodos de *clusters*, con estratos definidos por su Energía Libre, que proporciona una ilustración concisa de las relaciones geométricas y energéticas entre las estructuras muestreadas en la trayectoria (ver Sección IV.3.1). Estas estructuras de nodos están ejemplificadas en la Figura III.2.

Espacio para anotaciones

PARTE II

Implementación y resultados

IV.1 Construcción de los modelos computacionales 79
IV.2 Simulación del complejo en distintos estados de ocupación 80
IV.3 Comportamiento de RsmZ y de los complejos RsmZ-RsmE 81
IV.3.1 Clusterización y caracterización estructural 87
IV.3.2 Caracterización estructural de RsmZ y sus complejos con RsmE 91
IV.3.3 Accesibilidad de los motivos de unión en RsmZ 93
IV.4 Conclusiones del capítulo 102



IV. Sobre el mecanismo de captura RsmE de por RsmZ

Esta etapa de investigación se llevó a cabo entre los años 2018 y 2020. Parte de los resultados presentados en este capítulo fueron publicados en "The Journal of Physical Chemistry B" en el año 2021 [2].

Resumen de este capítulo La acción de RsmE de *Pseudomonas protegens* es regulada por los ARNp RsmX, RsmY y RsmZ, los cuales restauran la traducción de los genes reprimidos por la proteína mediante su captura. Así, estas moléculas actúan en tándem como un sistema regulador de la expresión génica. En este capítulo se presenta un análisis basado en simulaciones de DM que explora la diversidad conformacional de RsmZ y de los complejos RsmE-RsmZ. Los resultados revelan un patrón definido de exposición secuencial de los diferentes motivos de unión de RsmZ causado por la captura de las sucesivas unidades de RsmE. En conjunto, estas observaciones brindan una explicación simple y consistente para el orden de unión observado experimentalmente, y cuyas bases moleculares no habían sido develadas hasta el momento. A su vez, se propone una posible explicación para las discrepancias presentadas por las referencias 73 y

78 respecto a cuál es el orden de unión tras ocuparse el primer bolsillo de RsmZ. Este capítulo, además, presenta modelos estructurales para los complejos RsmE-RsmZ parcialmente ocupados que no habían sido descritos de forma experimental al momento de escribirse esta tesis.

Las estructuras utilizadas como punto de partida en este capítulo son las reportadas en la Referencia 73. Éstas contienen a los primeros 72 nucleótidos de RsmZ, el cual se presenta en dos conformaciones denominadas L y R. Las estructuras secundaria y terciarias de ambas conformaciones están representadas en la Figura I.3 del Capítulo I. En el mismo estudio, se determinó que tanto para la conformación L como la R, la unión de RsmE a RsmZ se produce de forma cooperativa siguiendo un orden preciso, de manera que el ingreso del primer dímero de la proteína favorece el acceso del segundo, y así sucesivamente hasta que se ocupa la totalidad de los motivos de unión del ARNp. La secuencia de unión está indicada en la Figura I.4. La primera unidad de RsmE se une a los *stem-loops* 2 y 3, mientras que la segunda unidad es capturada por los *stem-loops* 1 y 4. El último motivo de este fragmento de RsmZ en ser ocupado es el segmento de simple cadena ubicado entre los *stem-loops* 2 y 3 [73].

En conjunto, los motivos formados por *stem-loops* permiten la unión de dos moléculas de RsmE, mientras el segmento simple cadena constituye la mitad del bolsillo de unión para una tercera. En otras palabras, uno de los sitios del tercer dímero de RsmE se une a la región monocatenaria del fragmento de RsmZ aquí simulado, mientras el sitio restante se une a un motivo situado más allá del nucleótido 72, cuya estructura no fue resuelta por la Referencia 73. Así, la molécula completa de RsmZ posee la capacidad de unir adicionalmente otros dos dímeros de RsmE en sitios que no se encuentran presentes en las estructuras estudiadas en este trabajo.

Las razones moleculares por las cuales el acceso de RsmE a RsmZ se produce en el orden antes mencionado no pudieron ser establecidas en el trabajo de 2014, aunque algunas hipótesis fueron entonces planteadas. Nuestras primeras simulaciones sobre el sistema RsmE-RsmZ tuvieron como finalidad explicar este fenómeno, y son descritas en el presente capítulo.

Hipótesis trabajadas en este capítulo
La hipótesis que originó el trabajo descrito en este capítulo es que **la dinámica con la que RsmZ recorre su diversidad conformacional permite explicar el orden en el que RsmE ocupa sus sitios de unión**. Respecto al carácter cooperativo de esta unión, la principal hipótesis fue que **la incorporación de cada dímero de RsmE induce cambios estructurales en RsmZ que favorecen el ingreso de la siguiente unidad proteica**. En ambos casos se asumió que los fenómenos postulados eran susceptibles de ser analizados por técnicas de DM (Ver Secciones II.1.3 y III.3.1).

Objetivos de este capítulo
Generales
- Contribuir al conocimiento integral de los mecanismos de regulación post- transcripcional mediados por la familia de proteínas Csr/Rsm.
- Aplicar metodologías novedosas para el análisis de complejos proteína-ARN.

Específicos
- Explicar el carácter cooperativo y secuencial de la unión de RsmE a RsmZ.
- Comprender las bases moleculares para la selectividad del primer bolsillo de unión de RsmZ para RsmE.
- Caracterizar el comportamiento dinámico y estructural de RsmZ en presencia de diferentes unidades de RsmE.
- Observar si existen instancias de interconversión entre las conformaciones L y R.

En las siguientes secciones se describirán los pasos seguidos para abordar los objetivos de este capítulo, así como las técnicas utilizadas para analizar los resultados. Para abreviar, a partir de ahora me referiré a los motivos de unión de RsmZ con la sigla “MU”, mientras que los sitios de RsmE involucrados en estas uniones serán abreviados como “SU”. Los *stem-loops* 1, 2, 3 y 4 serán nombrados como SL1, SL2, SL3 y SL4, respectivamente, mientras el segmento de simple cadena entre SL2 y SL3 será mencionado como SC. Finalmente, utilizaré la nomenclatura SLX_{GGA} (con X entre 1 y 4) para referirme específicamente a la secuencia GGA de un SL particular. Todas estas siglas están resumidas en la Tabla VII.1.

IV.1 Construcción de los modelos computacionales

Los modelos computacionales de RsmZ y de los complejos RsmZ-RsmE fueron directamente generados a partir de las estructuras disponibles al momento de realizarse esta etapa de investigación [73]. Los códigos de entrada PDB son 2MF0 y 2MF1 para los confórmeros L y R, respectivamente.

Los modelos computacionales relacionados a estas estructuras PDB serán nombrados como L3 y R3, dado que corresponden a los confórmeros L y R, y contienen tres moléculas de RsmE cada uno. A partir de los mismos, se construyeron los modelos L2 y R2 removiendo la molécula de RsmE unida al segmento SC. Para esto, se eliminaron las coordenadas correspondientes a la tercera unidad de RsmE del archivo PDB de los modelos completos. Luego, se obtuvieron los modelos L1 y R1 eliminando la unidad de RsmE unida a $SL1_{GGA}$ y $SL4_{GGA}$ de L2 y R2. Por último, se generaron los modelos L0 y R0, correspondientes a RsmZ en estado libre, eliminando la última unidad de RsmE, unida a $SL2_{GGA}$ y $SL3_{GGA}$. La Figura IV.1 muestra una representación pictórica del procedimiento seguido. Nótese que el orden de salida de las unidades de RsmE en la figura es el opuesto al orden de unión descrito en la referencia 73.

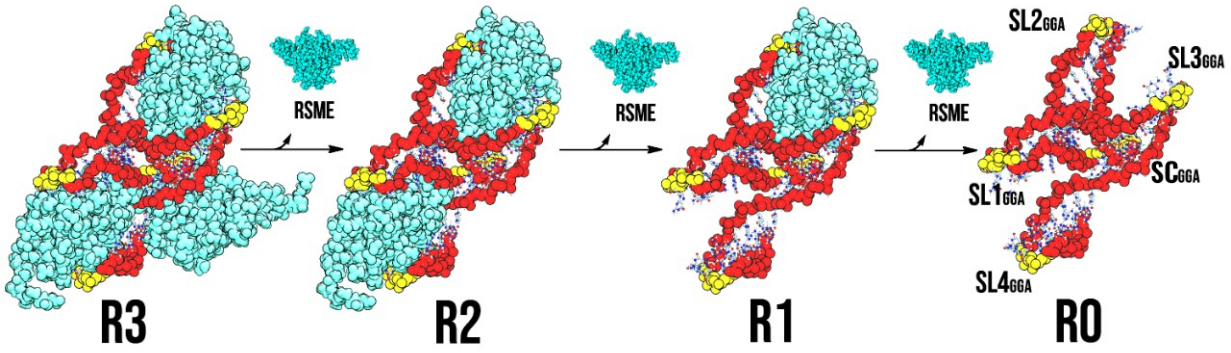


Figura IV.1: Representación esquemática de los modelos estudiados en este trabajo. Sólo la conformación R fue graficada. Un esquema similar se utilizó para el confórmero L. Los dímeros de RsmE están representados en celeste, mientras el esqueleto de RsmZ está mayormente en rojo, con excepción de sus MU, los cuáles están representados en amarillo. La imagen fue adaptada de la Referencia 2.

IV.2 Simulación del complejo en distintos estados de ocupación

El siguiente protocolo fue aplicado para realizar simulaciones de DM de todos los modelos del complejo RsmE-RsmZ generados como se explicó en la sección precedente. Los modelos fueron solvatados utilizando el módulo LEAP de AMBER [233]. Para esto se añadieron moléculas de agua en una caja octaédrica con paredes fijas a una distancia de 12.0 Å del átomo más próximo del complejo. A continuación, se añadieron iones Na^+ para neutralizar el sistema y luego se agregaron tanto iones Na^+ como Cl^- en cantidades equivalentes, hasta alcanzar una concentración 0.15M. Las interacciones entre las partículas del sistema se modelaron con el campo de fuerza ff14SB para las moléculas proteicas [298], el modelo TIP3P para las moléculas de agua [271], y la actualización ff99 bsc0$_{\chi OL3}$ para RsmZ [261] (Ver Sección III.3.1). Para los iones se utilizó la parametrización frcmod.ionsjc_TIP3P [299]. Las simulaciones fueron realizadas con los módulos SANDER y PMEMD de AMBER 18 [233, 300].

El sistema generado con el módulo LEAP fue primero minimizado y luego calentado durante 2 ns a volumen constante (NVT), hasta alcanzar una temperatura de 303K. La temperatura fue controlada con el termostato de Langevin con una frecuencia de colisión de 1.0 ps^{-1} [301]. Las longitudes de los enlaces que involucran átomos de hidrógeno fueron restringidas a través del algoritmo SHAKE. Acorde a esto, el paso de integración fue fijado en 2.0 fs (Ver Sección III.2). El método PME, con un radio de corte de 9.0 Å, fue implementado para las interacciones electrostáticas [233], por lo que las mismas fueron evaluadas en el espacio directo para $r < 9$Å y en el espacio recíproco para $r >$9.0Å [251, 252]. Un valor de corte de 9.0 Å fue también utilizado para el resto de las interacciones no ligadas.

Luego de finalizar el calentamiento, se cambiaron las condiciones de NVT a NPT, permitiendo que el sistema relaje su densidad durante 1ns. La presión fue controlada con el barostato de Berendsen con una constante de acoplamiento de 2.0 ps. Con la conformación alcanzada al terminar esta equilibración, se iniciaron 10 simulaciones de DM de 200 ns en condiciones NPT para cada modelo, alcanzando un total de 2.00 μs. Es importante destacar que ésto no equivale a una única trayectoria de 2.00 μs, sino que en cada simulación se utilizaron diferentes velocidades iniciales,

las cuales fueron tomadas al azar de una distribución de Maxwell con el fin de reproducir diversas regiones del espacio conformacional (Ver Sección III.3.2). Las estructuras adoptadas por el sistema durante las trayectorias fueron registradas cada 0.25ns, generando así una muestra total de 8 mil *snapshots* por modelo.

IV.3 Comportamiento de RsmZ y de los complejos RsmZ-RsmE

Para comprobar la estabilidad de los modelos y determinar el punto a partir del cual se encuentran equilibrados, se analizó el perfil de RMSD entre las estructuras muestreadas en la simulación y su correspondiente estructura inicial. La medición se realizó considerando sólo los átomos de Fósforo del esqueleto de RsmZ. La Figura IV.2 muestra ejemplos típicos de la evolución de los RMSD de cada modelo. Basándonos en el comportamiento de estas curvas, los primeros 150 ns de cada trayectoria se consideraron como un período de equilibración. De esta manera, se definió que la etapa de producción para cada sistema consiste de 10 trayectorias de 50 ns. Dado que las estructuras se registraron cada 0.25 ns, se obtuvo un conjunto de 2000 snapshots para cada uno de los 8 modelos.

Para cada uno de estos *sets* de datos, se calculó la FDP de su RMSD. La Figura IV.3 muestra este resultado. Se observa que todas las distribuciones son multimodales, presentando varios picos. Además, se vuelven más dispersas y alcanzan mayores valores de RMSD conforme menor es la cantidad de unidades de RsmE unidas a RsmZ. Así, los modelos L0 y R0 son los que presentan las distribuciones más anchas y alcanzan los mayores valores de RMSD. Esa tendencia se mitiga en los modelos con una y dos unidades de RsmE unidas, y alcanza su mayor rigidez en los modelos L3 y R3.

La forma en la que están computados los valores de RMSD representa una medida de cuánto se desvían las conformaciones muestreadas de RsmZ respecto a su estructura de partida. Intuitivamente, se puede pensar que esta característica está correlacionada con la flexibilidad de RsmZ, la cual aumenta conforme menos unidades de RsmE se encuentran formando el complejo, o bien que presenta estructuras de equilibrio progresivamente más diferentes a las de los complejos completamente ocupados. Sin embargo, los valores de este RSMD no permiten

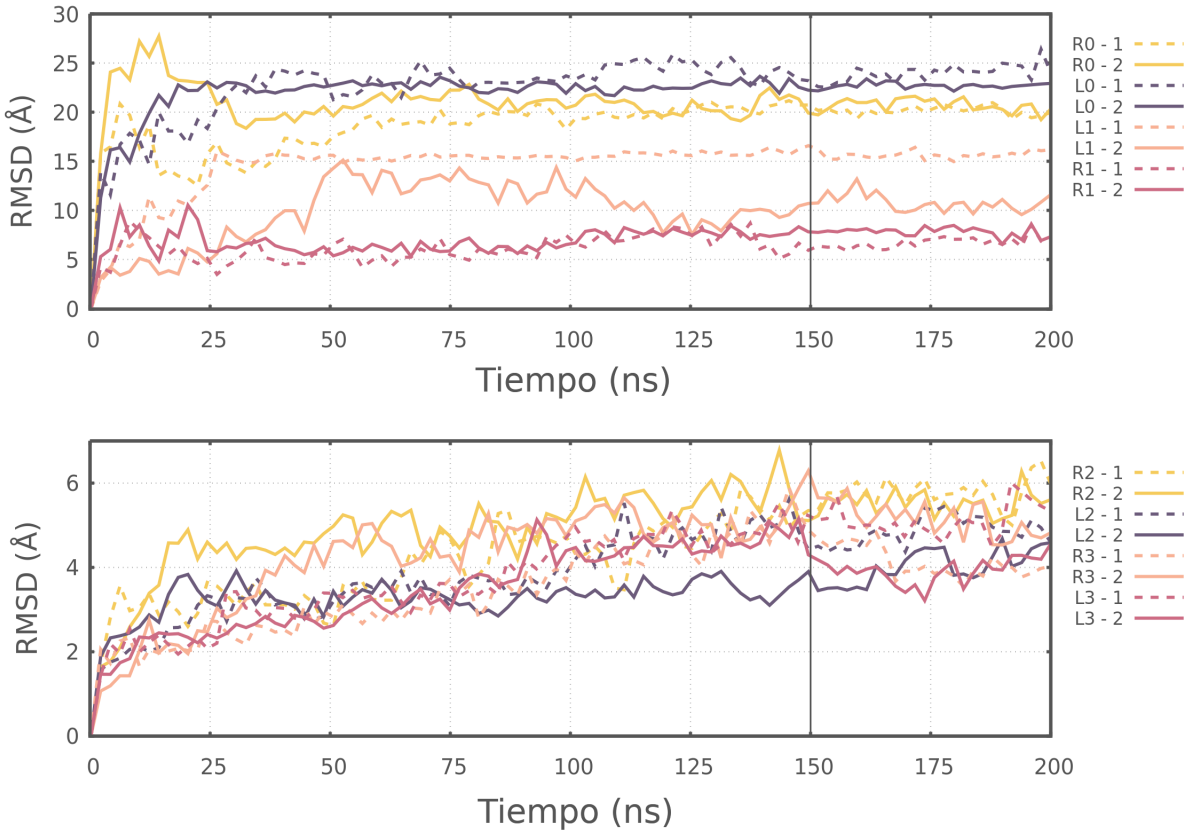


Figura IV.2: Evolución temporal del RMSD de los átomos de fósforo de RsmZ respecto de la estructura inicial. Se muestran los resultados de las simulaciones representativas de cada modelo. Con una línea gris se señala el punto a partir del cual las simulaciones alcanzaron valores estables de RMSD. La imagen fue adaptada de la Referencia 2.

evaluar cuánto se diferencian las estructuras entre sí. Para explorar este aspecto, se aplicó el método de PCA utilizando las mismas coordenadas que las empleadas para calcular los perfiles de RMSD.

Todas las simulaciones de cada modelo fueron concatenadas para construir la trayectoria que luego sería utilizada en el PCA. Se ha demostrado que este procedimiento mejora sistemáticamente la consistencia de los autovectores del PCA [289, 302]. Finalmente, las muestras de cada trayectoria fueron proyectadas sobre los primeros dos autovectores del PCA para obtener los componentes principales 1

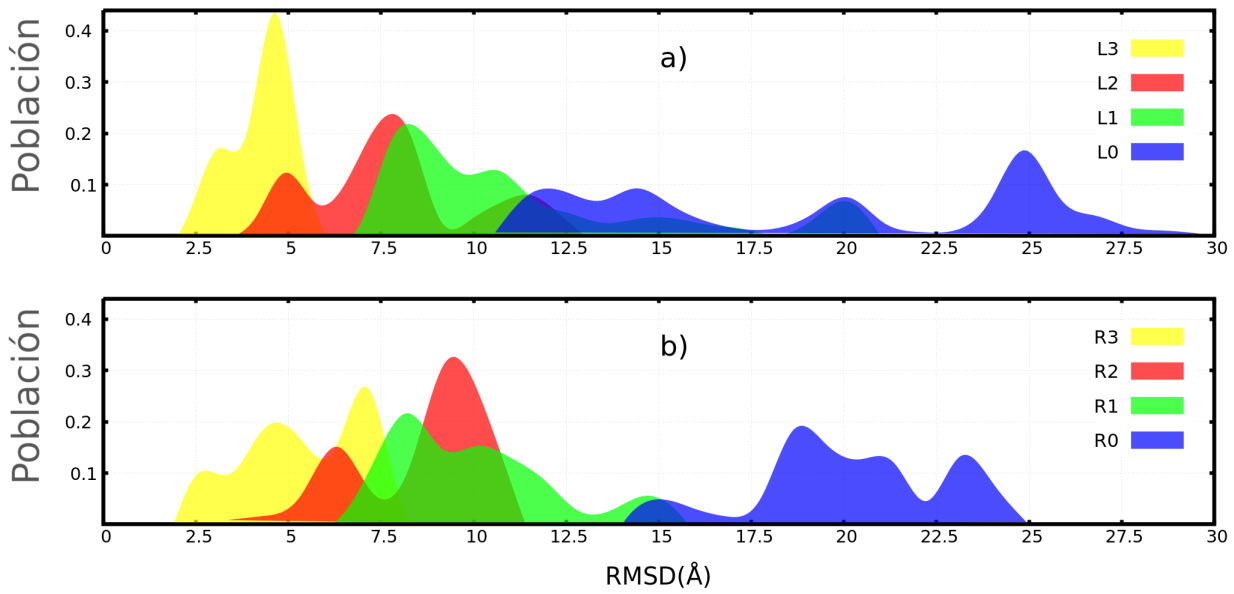


Figura IV.3: FDP para los RMSD de los átomos de P de RsmZ respecto a la primera estructura de cada modelo. Las curvas representadas en amarillo, rojo, verde y azul, corresponden a los modelos con 3, 2, 1 y ningún dímero de RsmE unido a las conformaciones L (Panel A) y R (Panel B) de RsmZ, respectivamente. La imagen fue adaptada de la Referencia 2.

y 2 (PC1 Y PC2) (Ver Sección III.5.1). Esas proyecciones se muestran en la Figura IV.4, y permiten tener una idea de cuánto difieren entre sí las estructuras de cada modelo.

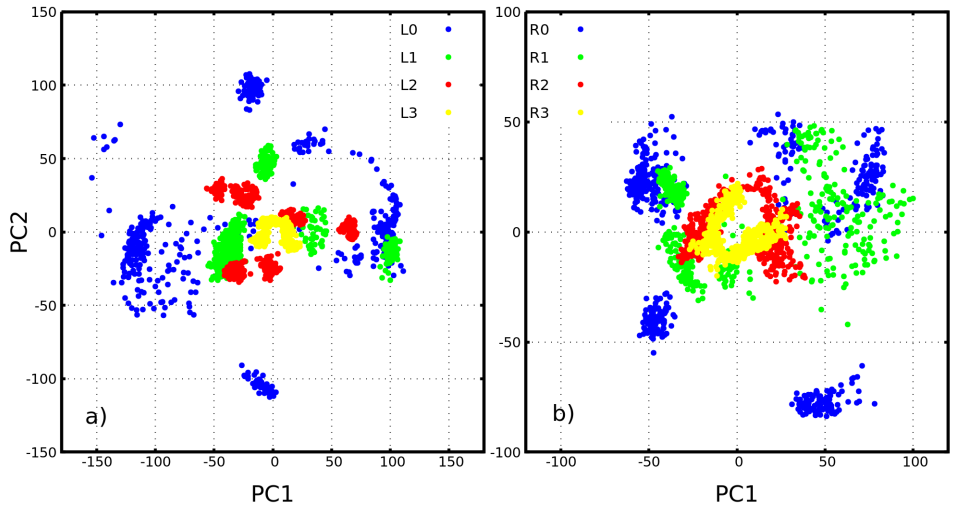


Figura IV.4: Proyección de las trayectorias sobre los primeros dos modos obtenidos por PCA. La información obtenida para los distintos modelos está representada en colores diferentes. Así, la distribución azul corresponde al modelo L0/R0, la verde al modelo L1/R1, la roja al modelo L2/R2 y la amarilla al modelo L3/R3. **Panel A**: muestra los resultados para la conformación L. **Panel B**: muestra los resultados para la conformación R. La imagen fue adaptada de la Referencia 2.

Una primera observación indica que las estructuras obtenidas a partir de las simulaciones de L3 y R3 ocupan una región relativamente pequeña de su respectivo subespacio, ubicada cerca del origen del plano formado por los dos primeros componentes principales. Sin embargo, los modelos con menor ocupación de RsmZ por parte de RsmE se alejan progresivamente del origen, formando distribuciones más dispersas de sus respectivos subespacios.

En la figura también es posible observar cómo las formas libres de RsmZ son particularmente flexibles, y que los modelos correspondientes a la conformación L tienen una mayor flexibilidad respecto al confórmero R. Una medida cuantitativa de esta observación está dada por los dos primeros autovalores para cada conformación. Para L0 estos son 7811 $Å^2$ y 3440 $Å^2$, respectivamente, mientras que los correspondientes a R0 son 2889 $Å^2$ y 2131 $Å^2$. Si contrastamos estos valores con los obtenidos a partir de los modelos con 3 proteínas, se observa un marcado

descenso a 114 Å^2 y 40 Å^2 para el caso de L3, así como a 151 Å^2 y 85 Å^2 para el caso de R3.

Tanto el análisis de RMSD como el método de PCA basado en coordenadas cartesianas requieren poder discriminar entre las deformaciones internas de la molécula y su rotación global. Sin embargo, dicha separación no es posible cuando las deformaciones moleculares son de gran magnitud [289]. Esta limitación se puede evitar utilizando coordenadas internas tales como distancias de contactos o ángulos diedros. En particular, se ha demostrado que el uso de ángulos diedros es apropiado para describir los cambios conformacionales en moléculas de ARN [296]. Este método, conocido como dPCA, fue aplicado para analizar las trayectorias de los modelos estudiados en este capítulo, a fin de generar coordenadas apropiadas para la clusterización de las estructuras muestreadas.

Los ángulos diedros utilizados para este segundo PCA fueron α, β, γ, δ, ε y ζ, los cuales describen la conformación del esqueleto del ARN (ver Figura IV.5). Además, en el análisis se incluyeron sólo los nucleótidos 16-19, 36-44, y 56-59, como se muestra en la Figura IV.5. Estos residuos corresponden a las regiones que empalman los cuatro SL del segmento considerado de RsmZ. El motivo de la elección de estos segmentos es que concentran las principales diferencias entre las conformaciones L y R [303].

Para calcular la matriz de covarianza del dPCA, se concatenaron todas las trayectorias de los modelos L y R correspondientes al mismo nivel de ocupación. Es decir, se concatenó a L2 con R2, a L1 con R1, y a L0 con R0. El objetivo de este procedimiento es poder usar el dPCA para analizar si en algún caso se produce un solapamiento entre las estructuras generadas comenzando en ambas conformaciones. El análisis se completa proyectando las trayectorias concatenadas sobre el primer autovector y luego calculando la FDP de los valores obtenidos. La Figura IV.6 muestra el resultado de dicho análisis. Así, se muestran las distribuciones de probabilidad para los valores correspondientes al primer componente principal.

Se observa que las distribuciones para los confórmeros de L y R en todos los complejos RsmZ-RsmE están separadas entre sí, mientras que presentan un mínimo solapamiento en los modelos sin RsmE. Sin embargo, incluso en ese caso, hay una

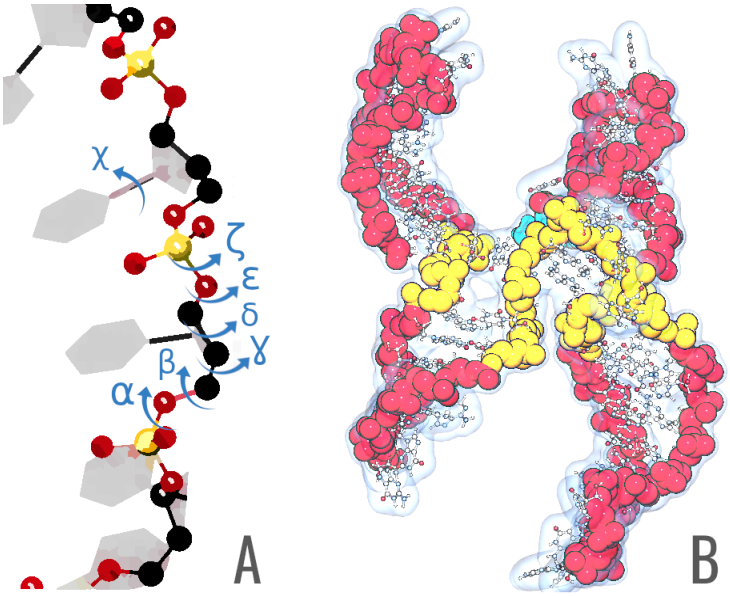


Figura IV.5: **Panel A**: Representación de los ángulos diedros de una molécula de ARN. Los átomos de Fósforo están representados en amarillo, mientras los de Oxígeno y Carbono se muestran en rojo y negro, respectivamente. Los ángulos, por su parte, están señalados en azul. Los ángulos α, β, γ, δ, ε y ζ corresponden al esqueleto del ARN, mientras el ángulo diedro χ corresponde a la base nucleotídica de la molécula, representada en gris. **Panel B**: Nucleótidos cuyos ángulos diedros fueron utilizados en la definición del dPCA para RsmZ. Los nucleótidos seleccionados están destacados en amarillo. El primer nucleótido de la secuencia está destacado en celeste. La imagen fue adaptada de la Referencia 2.

clara diferenciación entre las estructuras en las trayectorias iniciadas en L y las iniciadas en R, sugiriendo que no hubo interconversión entre ambas conformaciones durante las simulaciones. En este punto, cabe recordar la hipótesis planteada en la Referencia 73, que sugiere la importancia de que coexistan dos conformaciones de RsmZ para que éste pueda adaptar su mecanismo de captura a diferentes contextos estructurales.

La siguiente sección tiene por objetivo analizar las estructuras correspondientes a los mínimos de Energía Libre de cada modelo y racionalizar sus principales características estructurales. Para esto, se partirá del dPCA introducido en esta sección.

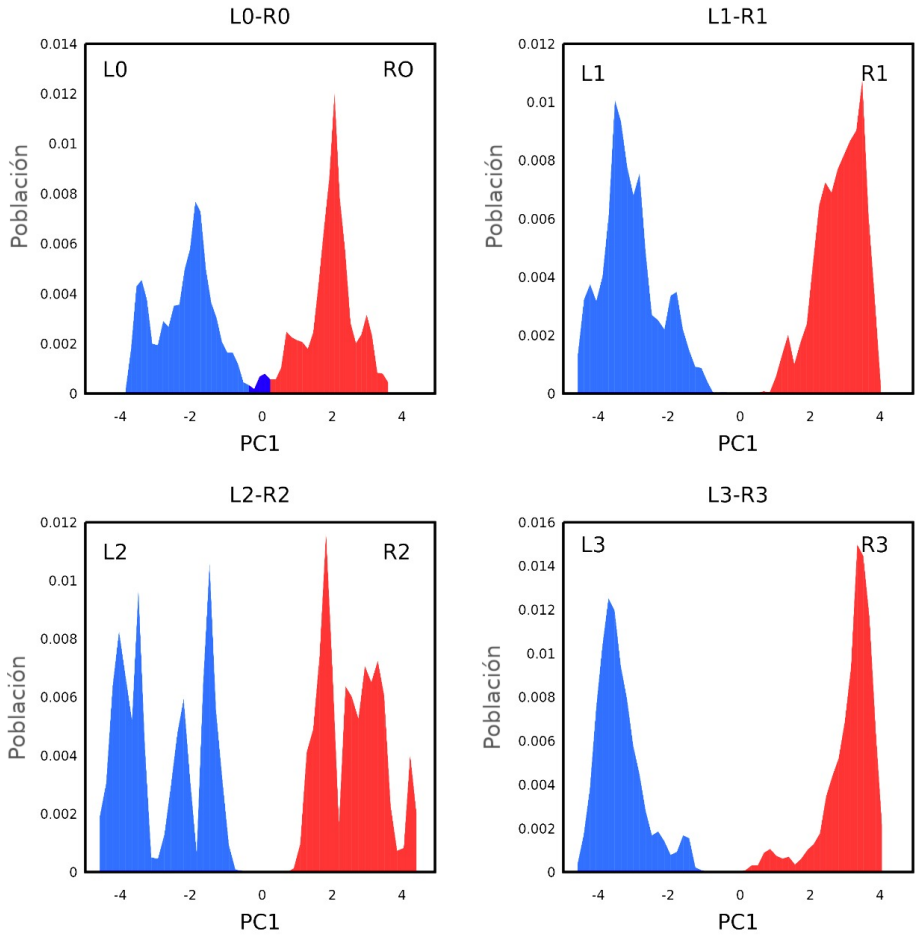


Figura IV.6: FDP para las proyecciones sobre el primer modo obtenido por dPCA. La distribución obtenida para la conformación L está presentada en azul, mientras la correspondiente a la conformación R está presentada en rojo. La imagen fue adaptada de la Referencia 2.

IV.3.1 Clusterización y caracterización estructural

Para identificar las estructuras típicamente adoptadas por RsmZ, tanto en su forma libre como con diferentes grados de ocupación por parte de RsmE, se implementó el método de clusterización basado en densidad de Stock y sus colaboradores [293] (Ver Sección III.5.2). Las estructuras fueron proyectadas sobre los espacios de dimensión reducida generados por los primeros 4 modos del dPCA.

Para seleccionar el tamaño del subespacio sobre el que se va a proyectar hay que tener en cuenta dos criterios contrapuestos. Por un lado, es preferible utilizar un subespacio de muchas dimensiones para no generar los llamados "errores de proyección". Estos refieren a una alteración en la conectividad de los mínimos debida a la información que se pierde con los grados de libertad que se descartan. Pero, por otro lado, es necesario tener una buena estadística del espacio considerado. Si para describir cada dimensión se requieren Np puntos de grilla y el número de dimensiones es Nd, entonces la cantidad total de puntos que deben ser incluidos en el análisis es Np^{Nd}. Este número crece muy rápidamente con Nd y, como la muestra es limitada, ocurre que muy rápidamente el cociente entre la cantidad de muestra y la cantidad de puntos de la grilla se vuelve muy chico, generando un aumento en el error estadístico. Para este trabajo se seleccionó un espacio de dimensión 4 (4D) porque ese es el máximo recomendable para la cantidad de muestra disponible [293]. El error de proyección relacionado con esta reducción del espacio no debería afectar al objetivo del análisis, dado que éste tuvo como meta identificar estructuras típicas para cada modelo, sin pretender discriminar entre situaciones más ambiguas.

Para cada punto, se estimó su densidad local contando el número de estructuras que se encontraban a menos de un cierto radio R_ρ en el subespacio 4D analizado. Luego, se seleccionó el valor de R_ρ computando las poblaciones para diferentes valores de este parámetro y analizando las distribuciones obtenidas. De esta manera, el R_ρ elegido fue el de mayor valor que presenta una caída exponencial en la población, como se observa en la Figura IV.7. Los valores de R_ρ elegidos oscilan entre 0.20 y 0.35, dependiendo del modelo.

Como se explicó anteriormente, este método de clusterización provee además una estimación para la Energía Libre de cada estructura muestreada (medida en unidades de k_BT). Cuanto mayor es la cantidad de puntos en torno a cada estructura considerada, menor será el valor de Energía Libre estimado para la misma. Así, el mínimo global de Energía Libre será asignado a la conformación con la mayor densidad de estructuras a su alrededor. No obstante, esta estimación no es suficiente para describir cómo se distribuyen estos puntos sobre la SEL de la molécula, sino que sólo asigna un valor relativo sin distinguir qué puntos con valores similares de

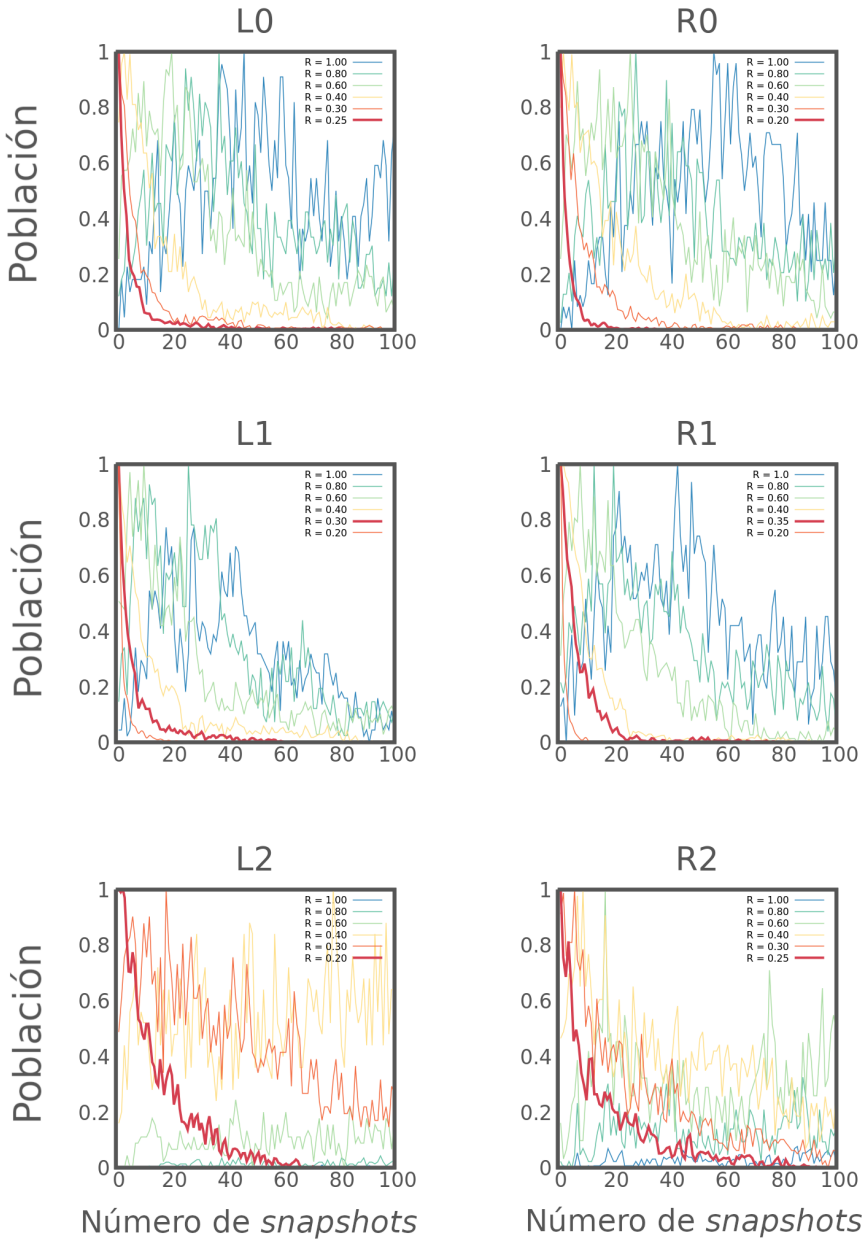


Figura IV.7: Efecto del radio (R_ρ) de la hiperesfera utilizada para estimar la Energía Libre local al realizar la clusterización por densidad. Las poblaciones están normalizadas para que el máximo sea igual a 1. Los valores seleccionados están destacados en rojo. La imagen fue adaptada de la Referencia 2.

energía se encuentran lejanos entre sí, y cuáles pertenecen al mismo mínimo local.

Para abordar este aspecto, se implementó la técnica de *lumping* propuesta en la Referencia 293 y discutida en la Sección III.5.2. Para esto se utilizó un paso de tamizado de 0.1 k_BT y un radio de clusterización igual a dos veces la distancia media del punto vecino más cercano. De este modo se construyeron los cladogramas que muestran cómo las estructuras se encuentran organizadas de acuerdo tanto a su proximidad en el subespacio proyectado como a su Energía Libre. En particular, las estructuras pertenecientes a los nodos terminales de cada árbol son las más representativas de cada subconjunto. Las estructuras típicas así obtenidas con cada uno de los modelos fueron tomadas de los nodos terminales de sus respectivos árboles. En la Figura IV.8 se muestran estos resultados.

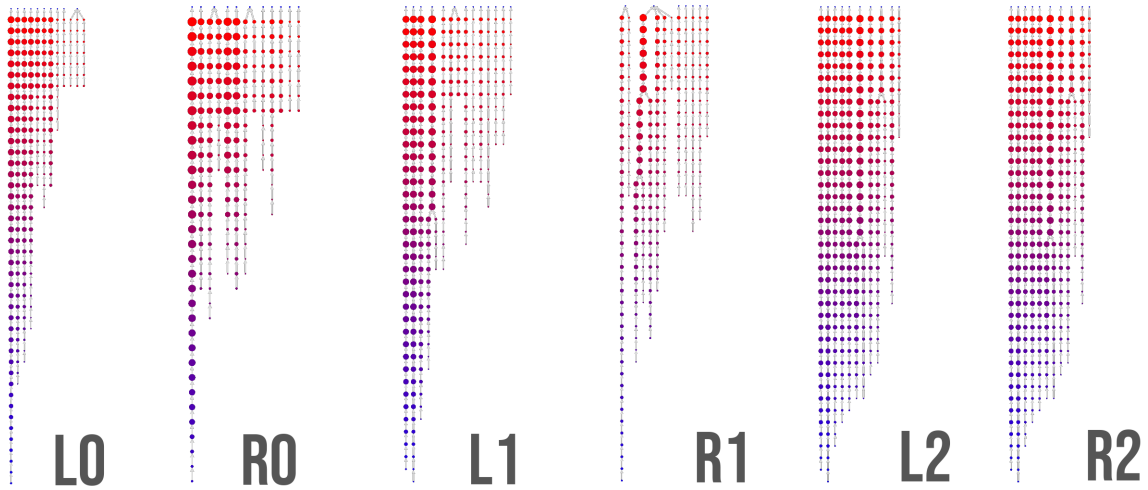


Figura IV.8: Cladogramas de clusterización basada en densidad. La longitud de cada rama corresponde a diferentes valores de Energía Libre, donde los nodos terminales de cada ramificación concentran a los mínimos de Energía libre de su respectiva conformación, y los valores máximos están representados en el margen superior del gráfico. Esta representación fue realizada utilizando el código disponible en `https://github.com/lettis`. La imagen fue adaptada de la Referencia 2.

IV.3.2 Caracterización estructural de RsmZ y sus complejos con RsmE

Al observar las estructuras más representativas de cada modelo, se notó un primer indicio acerca de la existencia de un patrón estructural que podría incidir sobre el orden seguido por RsmE para unirse a RsmZ. Se pudo notar que en los modelos L0 y R0, los únicos *stem-loops* que se encontraban a simple vista expuestos al solvente eran los correspondientes al primer bolsillo de unión. Los MU restantes aparecían ocluidos, interactuando entre sí o con su propia región doble cadena. En otras palabras, RsmE no podría acceder a los MU restantes. Por su parte, en los modelos L1 y R1, están disponibles los MU de los SL 1 y 4, pero no se observan estructuras en las que esté disponible el MU de la región SC. Sólo en los modelos L2 y R2 se observan estructuras en donde el segmento SC no se encuentra ocluido. Es interesante destacar que las interacciones que anulan la exposición de un determinado MU generalmente están mediadas por otros MU.

Para ilustrar estas observaciones, se seleccionaron dos estructuras representativas de cada estado de ocupación. En cada par, una estructura presenta al MU disponible para ser ocupado por RsmE, mientras que en la otra se encuentra ocluido. Las estructuras seleccionadas pueden visualizarse en `https://sketchfab.com/aormazabal`. La Figura IV.9 muestra una vista en perspectiva para estas estructuras, así como códigos QR que, al escanearse, permiten acceder a representaciones en Realidad Aumentada para cada modelo. A continuación, se describen las características estructurales de cada conformación elegida. Éstas serán nombradas de acuerdo a su estado de ocupación, explicitando la ubicación del MU disponible o indicando si todos sus MU se encuentran ocluidos.

<u>L0R0-SL3-accesible</u>. En este caso, los SL ocluidos anulan su disponibilidad entre sí o se encuentran flexionados sobre sí mismos. En particular, los MU de SL1 y SL4 interactúan entre sí, cancelando mutuamente su accesibilidad. La interacción se debe mayormente al apilamiento entre los residuos G9($SL1_{GGA}$) y C67($SL4_{GGA}$). Adicionalmente, interacciones de puente de Hidrógeno entre los residuos G9($SL1_{GGA}$) y G65($SL4_{GGA}$) contribuyen a ese comportamiento. Por su parte, los residuos U11, A12 y G13 de SL1 se ubican cerca de la región SC_{GGA},

inhibiendo su accesibilidad. El SL2 se encuentra flexionado sobre sí mismo, por lo que su MU interactúa con los nucleótidos ubicados en el origen de su región doble cadena, quedando impedido de capturar a RsmE. Finalmente, SL3 se encuentra abiertamente expuesto al solvente, pudiendo así capturar a la proteína.

L0R0-ocluido. En esta estructura, todos los MU están ocluidos debido a las interacciones de unos con otros. Las regiones $SL1_{GGA}$ y $SL2_{GGA}$ interactúan entre sí por apilamiento a través de los nucleótidos U11 y C30. Una situación análoga ocurre entre las regiones SC_{GGA} y $SL4_{GGA}$, que interactúan a través de los nucleótidos C42 y C67 de cada una de ellas. Finalmente, SL3 está ocluido debido a que su MU interactúa con los átomos de ribosa de U36, así como con la región simple cadena de SL2 y el residuo C37 ubicado en el inicio del segmento SC.

L1R1-SL1-accesible. En este caso, se puede observar una marcada diferencia en el comportamiento de los SL según si están unidos o no a RsmE. SL4 se encuentra ocluido, y el motivo GGA de su MU se encuentra cercano al SL3 (que junto al SL2 están ya ocupados por RsmE). Más aún, la región inicial de su región apareada se encuentra próxima al segmento SC_{GGA}, dificultando también el acceso de RsmE. Por otra parte, los residuos A41 y C42 de este MU, así como A43 y U57, interactúan entre sí formando un motivo del tipo *ribose zipper*. Contrariamente, SL1 se encuentra expuesto al solvente, con su motivo GGA disponible para unirse a RsmE.

L1R1-ocluido. En este caso, SL1 y SL4 están dispuestos de manera que sus respectivos motivos GGA quedan cercanos entre sí. Este comportamiento está estabilizado por las interacciones entre los residuos G9 y U11 ($SL1_{GGA}$) con G65 y G64 ($SL4_{GGA}$). Respecto a la región SC_{GGA}, éste se encuentra próximo a los segmentos de doble cadena de SL2 (U36) y SL4 (U58). Esta observación sugiere que no se encuentra disponible para su unión con RsmE. Los SL2 y SL3 se encuentran por su parte ocupados en la interacción con RsmE.

L2R2-SC-accesible y L2R2-ocluido. En estos modelos, los cuatro MU localizados en SL están interactuando con RsmE. Así, el motivo SC_{GGA} es el único que está desocupado. Debido a la estabilización estructural generada por la unión de dos unidades de RsmE, las estructuras observadas en L2 y R2 son similares a sus

contrapartes L3 y R3, excepto por el segmento SC que es el único que conserva aún cierto grado de flexibilidad. Esto le permite alternar entre estados de oclusión y disponibilidad. En las conformaciones seleccionadas para este par de modelos se puede apreciar claramente dicha modulación. En "L2R2-SC-accesible", las regiones apareadas de SL2 y SL4 se encuentran distantes del segmento SC_{GGA}. Además, hay un apareamiento de Wobble entre C37 y A41 en esta estructura. Debido a dicha interacción, los nucleótidos G_{38}, G_{39}, y A_{40} quedan expuestos al solvente. La interacción entre C37 y A41 no existe en "L2R2-ocluido", lo que provoca que el residuo central G39 de la región SC_{GGA} quede orientado hacia el centro de RsmZ. Por su parte, los nucleótidos C59 y C60 de SL4 se disponen próximos a SC_{GGA}, sugiriendo que queda inaccesible para RsmE.

Intuitivamente, estas estructuras sugieren que las conformaciones representativas de cada modelo oscilan entre estados de disponibilidad y no disponibilidad para la unión de RsmE. Este mecanismo sigue una secuencia de exposición concertada, coincidente con el orden en el que son ocupados los sitios de RsmZ según reportes experimentales [73]. La siguiente sección tiene por objetivo racionalizar y cuantificar las observaciones descritas para las estructuras típicas seleccionadas, e identificar si las disponibilidades de dichas conformaciones son sostenidas a lo largo de las trayectorias de DM.

IV.3.3 Accesibilidad de los motivos de unión en RsmZ

En la sección anterior se sugirió que el grado de exposición al solvente de los diferentes MU de RsmZ podría ser el determinante del orden de ocupación de dichos MU observado en el trabajo de Duss y sus colaboradores [73]. Para cuantificar ese grado de exposición, se definió la “accesibilidad" de cada motivo de la siguiente manera:

Se delimitaron cinco regiones 3D en RsmZ, cada una de ellas asociada con un MU determinado. Cada región contiene a todos los nucleótidos que, en la estructura PDB original, se encuentran dentro de una esfera de radio r_{ex} centrada en el centro de masa del GGA del MU considerado. Se considera que estas esferas definen una zona de exclusión para los nucleótidos que se encuentran en otras partes de

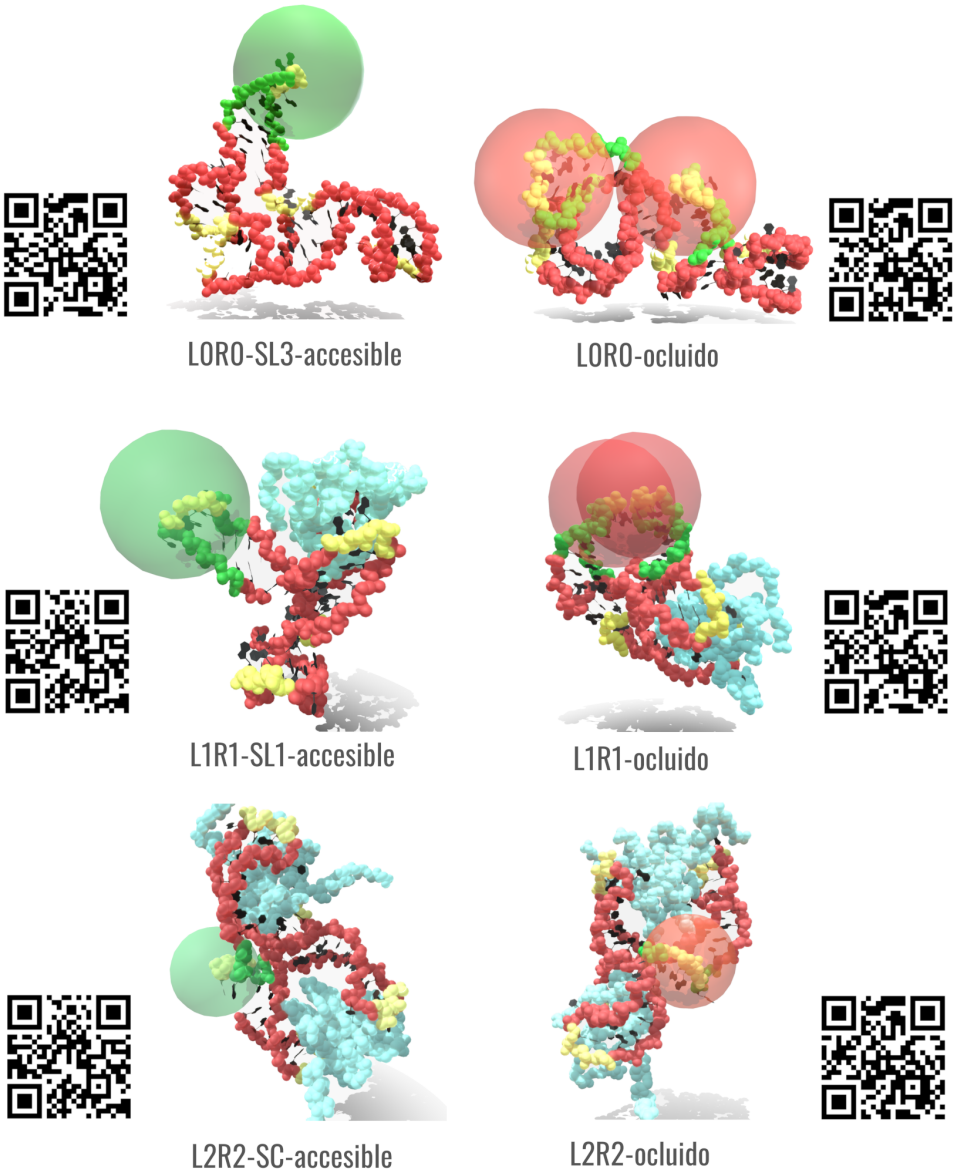


Figura IV.9: Vistas en perspectiva de las estructuras representativas para las conformaciones "L0R0-SL3-accesible", "L0R0-ocluido", "L1R1-SL1-accesible", "L1R1-ocluido", "L2R2-SC-accesible" y "L2R2-ocluido". En rojo se señala el esqueleto del ARN; en verde se representan las regiones doble-cadena de los SL estudiados; en amarillo se destacan los motivos GGA y en celeste se indica a RsmE. Las esferas verdes y rojas indican que un dado motivo está accesible u ocluido, respectivamente. Escaneando los códigos QR pueden visualizarse representaciones en Realidad Aumentada para cada modelo.

la molécula (es decir, los átomos que no son parte de esa región). Así, el GGA de cada MU se considera "accesible" cuando no hay átomos de otras regiones dentro de su esfera de exclusión. En caso contrario, se considera ocluido. Una representación animada de cómo funciona esta definición se presenta en el siguiente enlace: https://www.youtube.com/watch?v=_KwLNOe3jnI. En esa animación se muestra en color verde la situación en la que la esfera considerada no posee átomos de otras regiones en su interior, mientras el color rojo representa el caso en el que el MU se encuentra ocluido.

Es evidente que, con esta manera de estimar la accesibilidad, los resultados van a depender del radio de la esfera de exclusión. Sin embargo, para este trabajo, lo importante son las disponibilidades relativas de los sitios, y no así sus valores absolutos. El radio elegido para las esferas de exclusión fue de 18.0 Å para las centradas en los extremos de los SL y de 14.0 Å para el segmento SC, basándonos en los volúmenes necesarios para la ocupación de RsmE en las estructuras reportadas por la Referencia 73. La Figura IV.10 representa cómo quedan definidas dichas esferas para SL1 y la región SC. Además, para ver si las accesibilidades dependían del valor dado a r_{ex}, se hicieron pruebas con diferentes valores para este parámetro. Tal como se esperaba, se observó que su variación modifica las accesibilidades absolutas, pero prácticamente no cambia las relativas. La Figura IV.11 muestra los resultados de este ensayo.

Teniendo en cuenta los alcances y limitaciones de la definición utilizada, se procedió a determinar la accesibilidad de cada MU de RsmZ en todos los modelos. Para los modelos que contienen una o dos moléculas RsmE ya unidas, sólo se aplicó este análisis sobre los MU que no se encuentran previamente ocupados. Los porcentajes de accesibilidad estimados para cada MU se muestran en la Figura IV.12. Al examinar la accesibilidad de los cinco MU de RsmZ en sus diferentes estados de unión (X0 $\rightarrow$ X1 $\rightarrow$ X2 $\rightarrow$ X3), se confirma la presencia de un patrón estructural explica el orden de unión de RsmE.

En las formas libres de RsmZ, sólo $SL2_{GGA}$ y $SL3_{GGA}$ tienen porcentajes de accesibilidad significativos. Para L0, los porcentajes de disponibilidad alcanzan valores de 16.0% y 27.10%, respectivamente. En el caso de R0, los valores son 6.95%

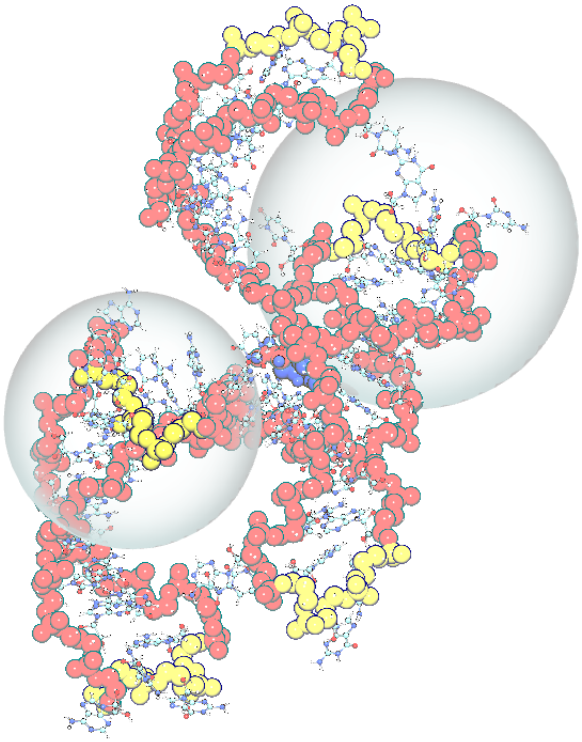


Figura IV.10: Esferas utilizadas para definir como accesibles a los motivos $SL1_{GGA}$ y SC_{GGA}. Se definió un radio de 18.0 Å para las esferas centradas en los extremos de los SL y de 14.0 Å para el que está en el segmento SC. La imagen fue adaptada de la Referencia 2.

y 18.85%. Los motivos restantes tienen accesibilidades marginales, o directamente nulas. Cuando se analizan los resultados de los modelos L1 y R1, se encuentra que el motivo SC_{GGA} permanece ocluido, pero se evidencia un marcado incremento en la accesibilidad de $SL1_{GGA}$ en ambas conformaciones, alcanzando valores de 21.35% en L1 y 19.70% en R1. En ésta última, hay también un aumento en la disponibilidad de $SL4_{GGA}$, que llega al 12.65%. La única excepción a esta tendencia ocurre en $SL4_{GGA}$ del modelo L1. Esto podría explicarse por una limitación del modelo, debido a que la estructura de RsmZ utilizada no está completa, sino que se encuentra truncada justo después de SL4. También podría ser un resultado genuino, aunque para corroborarlo sería necesario realizar el mismo análisis pero con un

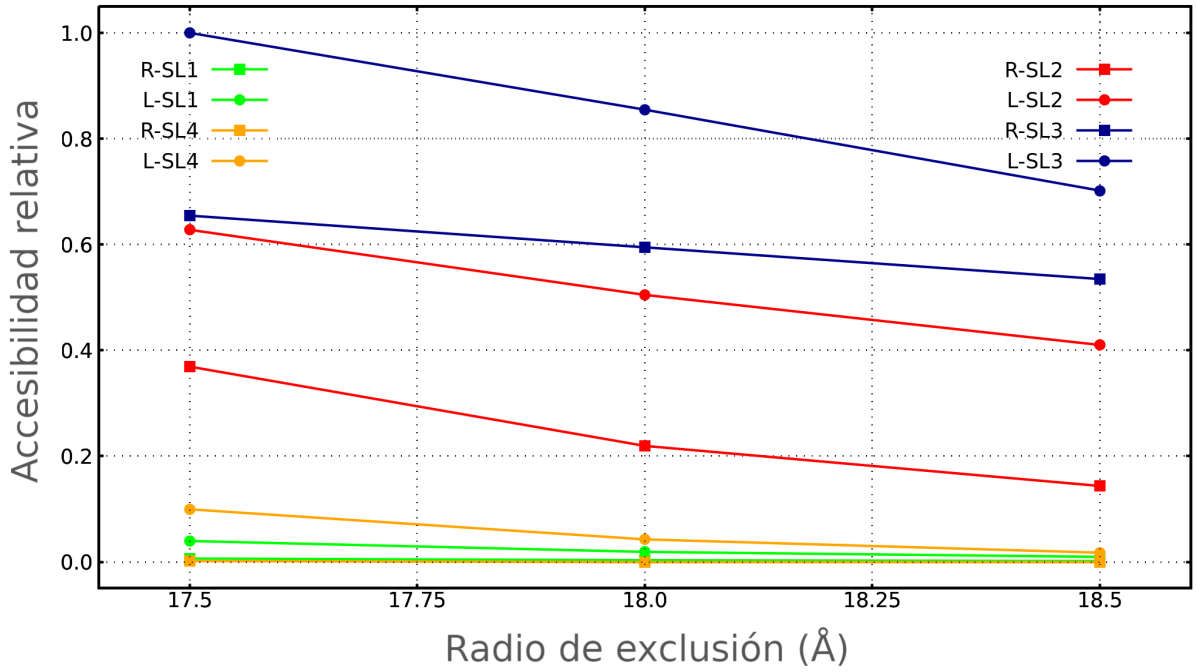


Figura IV.11: Sensibilidad del análisis de accesibilidad para los MU de RsmZ cuando se cambia el radio de exclusión. La imagen fue adaptada de la Referencia 2.

modelo sin la secuencia truncada.

Respecto a esto último, cabe destacar que poco tiempo después de concluirse este trabajo, se publicó un artículo en el que se reportó la estructura completa de RsmZ [78]. La misma contiene los 117 nucléotidos de la molécula, unidos a tres dímeros de RsmE, y se determinó el orden en el que se ocuparon los sitios de unión. Al igual que lo hallado con el fragmento de 72 nucleótidos estudiado en este capítulo, los autores indicaron que el primer dímero de RsmE se une a los MU de SL2 y SL3. Sin embargo, a diferencia de lo que ocurre con el fragmento de 72 nucleótidos, el segundo dímero se une a SL1 y no a SL4, sino a SL5 (un SL que se ubica más allá del nucleótido 72). Esto muestra que el orden de unión está sujeto a la extensión del ARNp. Nuestros resultados son coincidentes con el patrón observado en la Referencia 73, pero para explicar el comportamiento de RsmZ nativo este mismo

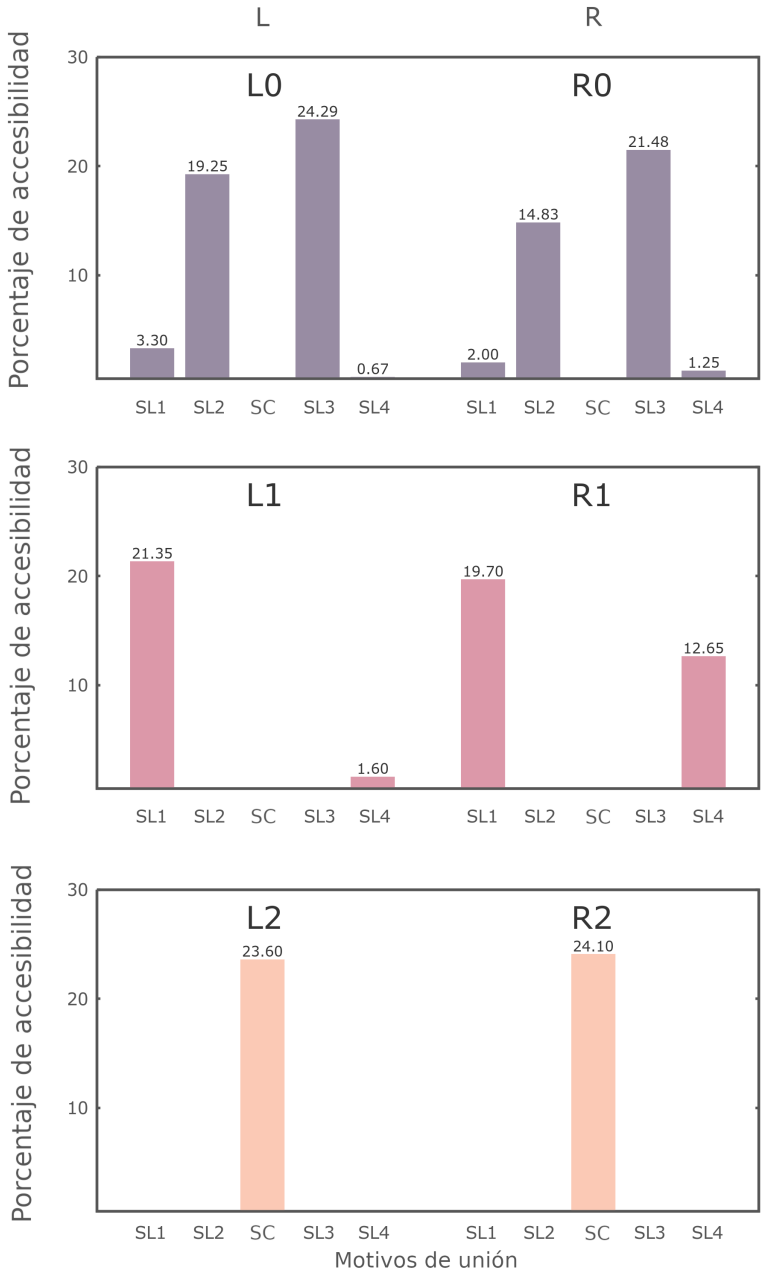


Figura IV.12: Porcentaje de disponibilidad para los distintos MU de cada modelo. Las barras violetas representan el resultado para L0 y R0, mientras las barras rojas corresponden a los modelos L1 y R1. Los modelos L2 y R2 están representados en naranja. Imagen adaptada de la Referencia 2.

análisis debería aplicarse a la molécula completa. No obstante, el comportamiento ambiguo observado para SL4 en el fragmento de 72 nucléotidos muestra que aún este constructo no está optimizado para utilizar SL4 para unir al segundo dímero de RsmE. Esto podría explicar la diferencia entre ambas referencias por ser éste el SL que difiere en el segundo bolsillo de unión de ambos modelos.

Completando el análisis, los únicos casos en los que se observa accesibilidad para el motivo SC_{GGA} son las simulaciones correspondientes a los modelos L2 y R2. Así, los porcentajes de disponibilidad pasan desde valores marginales a 23.60% y 24.10% para L2 y R2, respectivamente. Cabe destacar que estos son los mayores porcentajes de disponibilidad en toda la serie de datos.

L0	**SL1**$_{GGA}$	**SL2**$_{GGA}$	**SS**$_{GGA}$	**SL3**$_{GGA}$	**SL4**$_{GGA}$
SL1$_{GGA}$	0.00	0.00	0.00	0.00	0.00
SL2$_{GGA}$	-	**34.33**	0.00	1.84	0.80
SS$_{GGA}$	-	-	0.00	0.00	0.00
SL3$_{GGA}$	-	-	-	**60.69**	0.11
SL4$_{GGA}$	-	-	-	-	2.19

Tabla IV.1: Porcentajes de disponibilidad para los MU de L0. Los elementos diagonales se refieren a motivos individuales, mientras aquellos que están sobre la diagonal responden a combinaciones entre pares de motivos distintos.

Por último, se analizó con qué frecuencia dos MU se encuentran simultáneamente disponibles. Para ello, del conjunto de estructuras muestreadas para cada modelo, se seleccionaron las que tenían disponible al menos uno de los MU. Luego, sobre ese total, se observó en qué porcentaje estaba también disponible alguno de los MU restantes. Los resultados para L0 y R0 se encuentran disponibles en las tablas IV.1 y IV.2, respectivamente, mientras los de L1 y R1 están en las tablas IV.3 y IV.4. Los elementos diagonales de estas tablas muestran el porcentaje de estructuras que poseen un sólo MU disponible, mientras que los elementos por fuera de la diagonal indican los porcentajes en los que el par de MU indicados por la columna y la fila están simultáneamente disponibles.

R0	**SL1**$_{GGA}$	**SL2**$_{GGA}$	**SS**$_{GGA}$	**SL3**$_{GGA}$	**SL4**$_{GGA}$
SL1$_{GGA}$	0.00	0.00	0.00	0.00	0.00
SL2$_{GGA}$	-	**25.49**	0.00	1.98	0.00
SS$_{GGA}$	-	-	0.00	0.00	0.00
SL3$_{GGA}$	-	-	-	**72.53**	0.00
SL4$_{GGA}$	-	-	-	-	0.00

Tabla IV.2: Porcentajes de disponibilidad para los MU de R0. Los elementos diagonales se refieren a motivos individuales, mientras aquellos que están sobre la diagonal responden a combinaciones entre pares de motivos distintos.

La primera observación que surge al analizar estas tablas es que la mayor parte del tiempo, los SL que se encuentran expuestos lo hacen en soledad, es decir, sin que el resto de los MU también estén accesibles. En el caso de L0 y R0, si bien existen estructuras en las cuales existe la disponibilidad simultánea de dos

L1	$SL1_{GGA}$	SS_{GGA}	$SL4_{GGA}$
$SL1_{GGA}$	**93.03**	0.00	0.00
SS_{GGA}	-	0.00	0.00
$SL4_{GGA}$	-	-	**6.97**

Tabla IV.3: Porcentajes de disponibilidad para los MU de L1. Los elementos diagonales se refieren a motivos individuales, mientras aquellos que están sobre la diagonal responden a combinaciones entre distintos pares de motivos.

R1	$SL1_{GGA}$	SS_{GGA}	$SL4_{GGA}$
$SL1_{GGA}$	**60.89**	0.00	0.00
SS_{GGA}	-	0.00	0.00
$SL4_{GGA}$	-	-	**39.11**

Tabla IV.4: Porcentajes de disponibilidad para los MU de R1. Los elementos diagonales se refieren a motivos individuales, mientras aquellos que están sobre la diagonal responden a combinaciones de pares de motivos.

SL, los porcentajes correspondientes a dichas estructuras son marginales. De hecho, la mayor probabilidad de encontrar esa situación se observa para el par $SL2_{GGA}$/$SL3_{GGA}$, donde los SL se encuentran simultáneamente disponibles en un 1.84% y 1.98% de los casos para L0 y R0, respectivamente. Estos valores son marginales si se los compara con las probabilidades de observar sólo uno de estos SL disponible.

Estas observaciones sugieren que RsmE inicia su unión a RsmZ o a complejos RsmZ/RsmE siendo capturado inicialmente por un único SL disponible. Necesariamente, luego debe producirse un reacomodamiento intramolecular para permitir que el SL restante se una a la proteína por el otro lado. Estos resultados sugieren una nueva hipótesis, según la cual la exposición de los distintos SL ocurre individualmente para evitar el escenario en que dos de ellos interactúen simultáneamente con diferentes unidades de RsmE, lo que disminuiría la efectividad de unión y, por ende, la eficiencia del mecanismo de captura. Esta hipótesis está a su vez reforzada por el hecho de que MU individuales presentan constantes de disociación más elevadas cuando interactúan con RsmE de manera aislada respecto a cuando se encuentran en el contexto de la molécula completa [72, 73].

IV.4 Conclusiones del capítulo

En este capítulo se presentaron los resultados de las simulaciones de DM realizadas sobre modelos computacionales de RsmZ y de los complejos formados entre RsmZ y una, dos o tres moléculas de RsmE. Estas fueron las primeras simulaciones reportadas para el sistema RsmE-RsmZ. A partir de ellas se pudo observar que, acorde a lo esperable, la estructura de RsmZ se vuelve progresivamente menos flexible cuanto mayor es el número de moléculas de RsmE unidas. En particular, las formas libres de RsmZ presentan una enorme flexibilidad. No obstante, encontramos que aun en las conformaciones más flexibles (L0 y R0), las estructuras iniciadas desde L no muestrean el espacio de R, y viceversa. Esto muestra que la interconversión entre estos confórmeros, aún en ausencia de RsmE, tiene una barrera de Energía Libre demasiado alta para ser captada por las simulaciones realizadas en este trabajo.

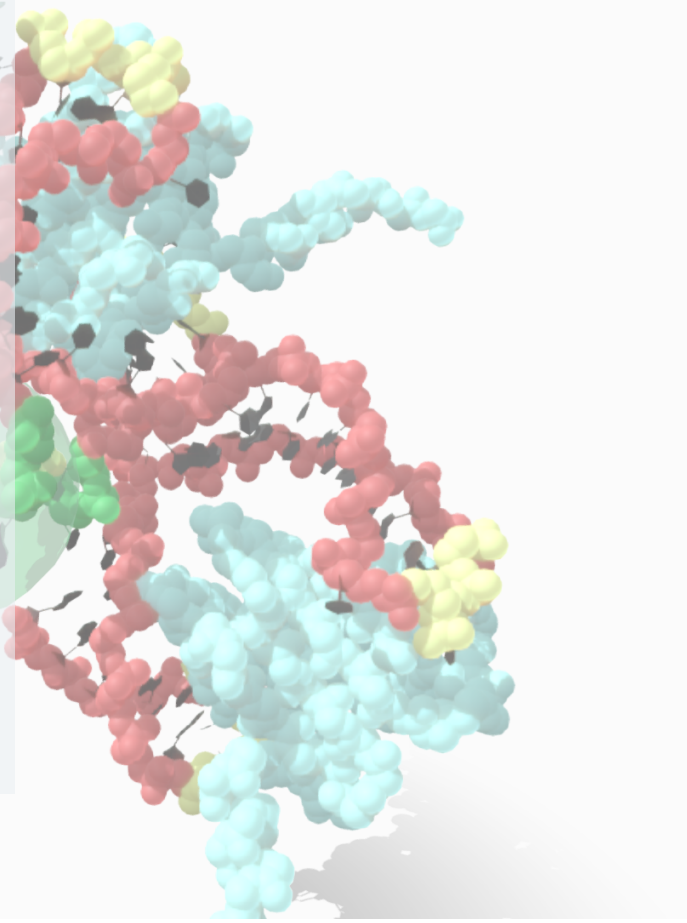
En cuanto al orden de ocupación seguido por RsmE, ambas conformaciones experimentan un comportamiento similar. Es destacable que su elevada diversidad conformacional no genera una mezcla de estructuras aleatorias. Por el contrario, cada modelo exhibe un patrón estructural consistente, en el cual los MU que deben unir al próximo dímero de RsmE se encuentran alternadamente expuestos mientras que el resto de los MU están ocluidos. Asimismo, las simulaciones revelaron que los SL que forman un mismo bolsillo de unión no están disponibles de manera simultánea, sino que se exponen al solvente de a uno por vez. Esto sugiere un mecanismo adicional de control para el sistema, dado que evita la posibilidad de que dos SL se unan a proteínas diferentes al mismo tiempo, lo que disminuiría su eficiencia de unión.

Finalmente, los resultados presentados en este capítulo sugieren la confirmación de las hipótesis planteadas al principio del mismo: Por un lado, fue posible develar un patrón estructural y dinámico que explica el carácter cooperativo para la unión de RsmE a RsmZ. Pero además, se sugiere que la unión de dicha proteína al ARNp ejercería una acción del tipo ajuste-inducido, en el cual las estructuras resultantes de ocupar cada sitio favorecen conformaciones en las que se habilita la ocupación del siguiente. Este punto será retomado en el capítulo siguiente. Habiendo también presentado adaptaciones de metodologías previas como técnicas de análisis diseñadas

ad hoc del modelo, se concluye que los resultados son satisfactorios respecto a los objetivos planteados inicialmente.

Espacio para anotaciones

V.1 Construcción de los modelos y simulaciones de DM estándar 108
V.2 Simulaciones de *Umbrella Sampling* 111
V.3 Mecanismo de unión entre RsmE y RsmZ 113
V.4 Cambios conformacionales durante el proceso de unión 118
V.4.1 Deformaciones de RsmE 118
V.4.2 Deformaciones de RsmZ-$(RsmE)_2$ 120
V.5 Deformaciones de RsmE y RsmZ-$(RsmE)_2$ aisladas 122
V.5.1 Análisis de Componentes Principales 125
V.5.2 Movimientos esenciales de RsmZ-$(RsmE)_2$ y RsmE 127
V.6 Conclusiones del capítulo 129
V.7 Anexo 132



V. Sobre el reconocimiento y unión de RsmE por RsmZ

Esta etapa de investigación se llevó a cabo entre los años 2020 y 2022. Parte de los resultados presentados en este capítulo, así como de los ensayos empleados para su obtención, fueron publicados en "Journal of Chemical Information and Modeling" en el año 2022 [3].

Resumen de este capítulo El Capítulo IV presentó simulaciones de DM que aportan una explicación para el orden seguido por RsmE al ocupar los diferentes sitios de unión del fragmento estudiado de RsmZ. En este punto, lo que resta es simular el evento de unión entre ambas moléculas. En este Capítulo se presentan simulaciones de *Umbrella Sampling* que describen el mecanismo de salida de RsmE del sitio de unión de RsmZ situado en la región simple cadena. El estudio está enfocado en dilucidar cuáles son las interacciones más relevantes del proceso. Los resultados proveen una estimación para la energía involucrada en el proceso de unión y muestran que el motivo estudiado de RsmZ adquiere una estructura símil *stem-loop* durante el evento. Cálculos adicionales revelaron que para poder llegar a esa conformación es necesario que el segmento

de simple cadena contenga al menos nueve nucleótidos. Finalmente, mediante un análisis de PCA de ambas moléculas en su forma libre, determinamos que sus principales coordenadas colectivas tienden a generar las conformaciones propicias para la unión. Los resultados sugieren fuertemente que la flexibilidad de la región monocatenaria de RsmZ afecta de manera crucial la capacidad de su MU para capturar a RsmE.

En este capítulo se simula un evento de unión entre RsmE y RsmZ. Para esto, se hizo uso de la técnica *Umbrella sampling* descrita en la Sección III.4. Con este método se simuló la salida de RsmE desde el MU de RsmZ situado en la región SC. Estas simulaciones se realizan en condiciones de cuasi-equilibrio, por lo que se asume que todos los grados de libertad del sistema, excepto el movimiento a lo largo de la coordenada de reacción, están equilibrados. Debido a este cuasi-equilibrio, el proceso observado es reversible. Por lo tanto, los datos colectados simulando la salida del dímero proveen información sobre el reconocimiento entre ambas moléculas y su posterior unión, cuando son analizados en sentido inverso.

Como se discutió anteriormente, el dímero de RsmE que se une a la región SC es el que ingresa en tercer lugar al constructo de 72 nucleótidos de RsmZ, y está unido por solo uno de sus lados. Esta situación provee un escenario favorable para estudiar el mecanismo de unión/separación entre la proteína y un motivo A(X)GGAX específico (que en este caso adopta la secuencia AGGAC), dado que la energía de esta unión es menor que la observada cuando RsmE es sostenido por ambos lados. Este hecho facilita la realización de cálculos de *Umbrella Sampling*, porque permite usar menores constantes de sesgo y, por ende, un menor número de ventanas.

En las siguientes secciones se describe cómo se obtuvieron los modelos utilizados para simular el proceso de separación del dímero de RsmE unido al motivo SC_{GGA} de RsmZ mediante simulaciones de *Umbrella Sampling*. Las siglas utilizadas en este capítulo están resumidas en la Tabla VII.1.

Hipótesis trabajadas en este capítulo
La principal hipótesis de este trabajo es que **la unión de RsmE al motivo SC_{GGA} de RsmZ induce cambios conformacionales en el ARNp** observables por simulaciones de DM. A su vez, se hipotetiza que **los residuos previamente descritos como críticos por su rol para la interacción entre RsmE y RsmZ presentan interacciones sostenidas durante la mayor parte del proceso de unión**.

Objetivos de este capítulo
Generales
- Aportar al conocimiento sobre el mecanismo de unión entre la familia de proteínas Csr/Rsm y los ARNp que ejercen una función de regulación sobre ellas.
- Analizar cómo influye la flexibilidad conformacional de los ARNp en los mecanismos de captura de RsmE mediante regiones situadas en empalmes.

Específicos
- Simular el proceso de unión entre RsmE y el motivo SC_{GGA} de RsmZ.
- Estimar el PEL del evento.
- Identificar los residuos claves para la interacción entre ambas moléculas, y la etapa del evento en la que participan.
- Caracterizar el comportamiento estructural y dinámico de las moléculas, tanto en sus formas libres como durante el proceso de unión.

V.1 Construcción de los modelos y simulaciones de DM estándar

Como se discutió en la Sección IV.3.3, el mecanismo de exposición al solvente del motivo SC_{GGA} es similar en ambas conformaciones de RsmZ. Así, tanto en la conformación L como en la R, este motivo se expone una vez que dos moléculas de RsmE ya se encuentran unidas a RsmZ mediante los *stem-loops* SL2/SL3 y SL1/SL4. Por otra parte, el análisis de las estructuras obtenidas del PDB indican que las interacciones entre RsmE y los motivos GGA en ambos confórmeros son equivalentes (ver Figura V.1). Sobre la base de estas observaciones, y habiendo corroborado que los cálculos de *Umbrella Sampling* realizados con ambas conformaciones arrojan resultados similares, en este capítulo se discuten los obtenidos con la conformación L.

La estructura correspondiente al complejo formado entre RsmZ y tres unidades de RsmE fue tomada del modelo L3 del Capítulo IV, pero en este capítulo se de-

nominará RsmZ-(RsmE)$_3$. Análogamente, el modelo correspondiente al complejo RsmE-RsmZ antes de la ocupación del tercer MU es el modelo L2 del Capítulo IV, pero en este capítulo se denominará como RsmZ-(RsmE)$_2$. El modelo correspondiente a RsmE en estado libre se obtuvo del archivo PDB de RsmZ-(RsmE)$_3$, tomando las coordenadas correspondientes a la unidad de RsmE que se une a RsmZ en tercer lugar.

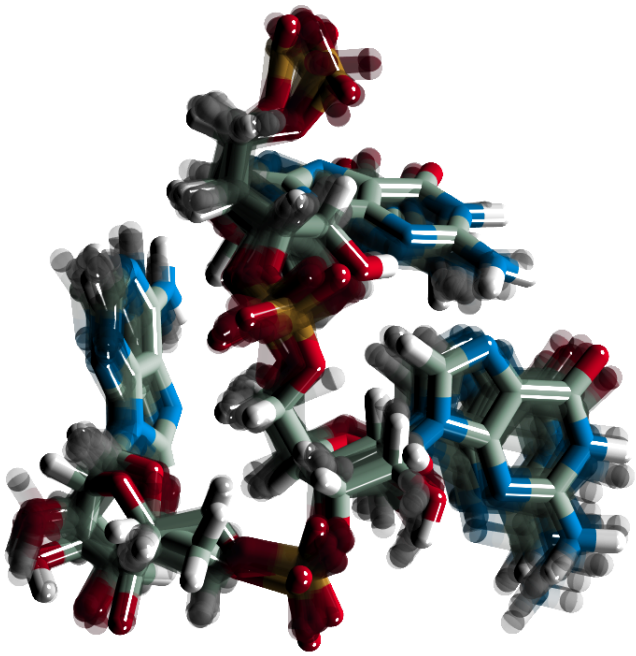


Figura V.1: Superposición entre los motivos GGA correspondientes a estructuras de RsmZ y *hcnA*, o partes de ellas, cuando dichos motivos se encuentran unidos a RsmE. En todos los casos, la disposición del motivo GGA presenta diferencias mínimas, de manera que la orientación adoptada al unirse a RsmE es prácticamente igual en todos los casos.

Finalmente, partiendo de RsmZ-(RsmE)$_2$ se construyeron otros dos modelos con modificaciones en la secuencia de simple cadena que contiene al motivo SC$_{GGA}$. En la estructura nativa, esta región se extiende desde el nucleótido U36 al U44 (ver Figura I.3). En uno de estos modelos "mutados", al que denominamos RsmZ^{m1}-(RsmE)$_2$, se eliminaron los nucleótidos U36 y U44 de la secuencia original. Simi-

larmente, para construir el modelo denominado RsmZ^{m2}-(RsmE)$_2$ se eliminaron C37 y A43 además de U36 y U44. Una versión esquemática de todos los modelos utilizados en este Capítulo se presenta en la Figura V.2.

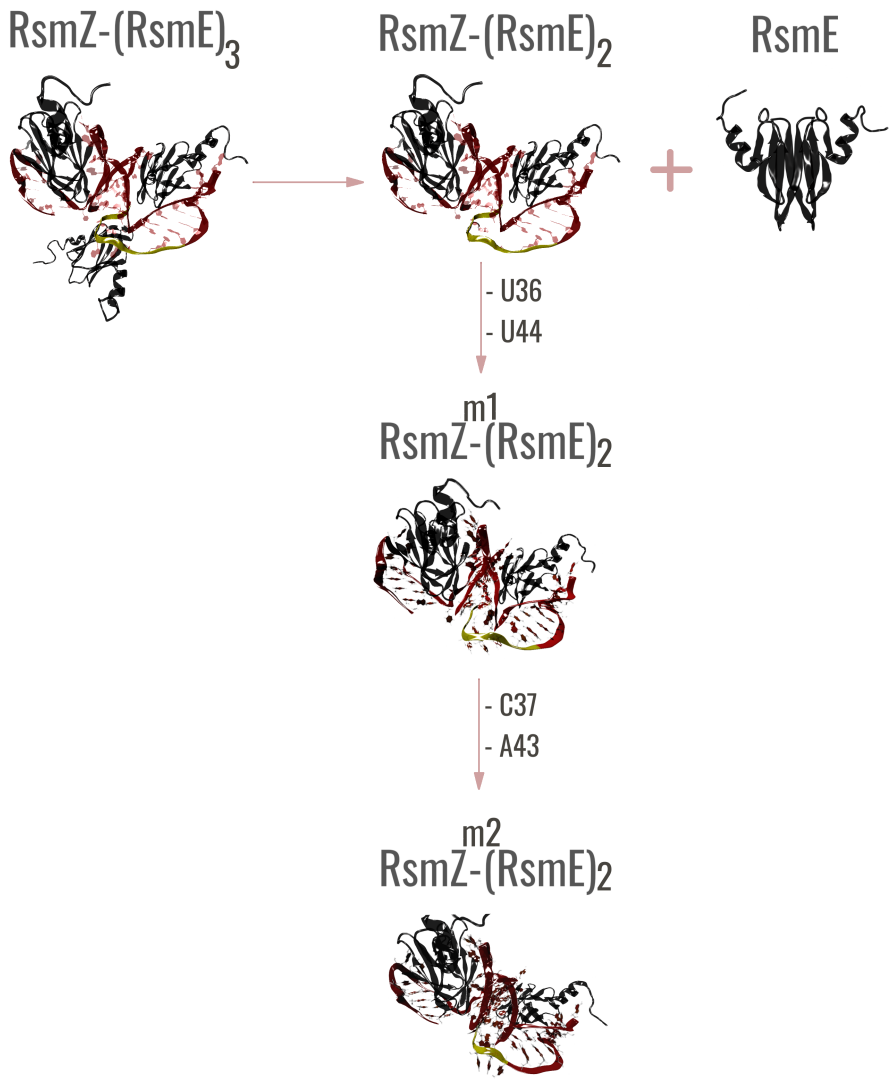


Figura V.2: Representación esquemática de los modelos estudiados en este capítulo. Se remarcan además los residuos que fueron sustraídos del segmento SC. Imagen adaptada de la Referencia 3.

Las simulaciones de DM de RsmZ-(RsmE)$_3$ y RsmZ-(RsmE)$_2$ fueron tomadas

del trabajo descrito en el Capítulo IV. Con el fin de establecer comparaciones significativas, las etapas de minimización, calentamiento, equilibración y producción de RsmZ^{m1}-(RsmE)$_2$ y RsmZ^{m2}-(RsmE)$_2$ se llevaron a cabo utilizando las mismas condiciones y parámetros que en dichas simulaciones. Lo mismo aplica para RsmE libre, con la única excepción de que, por alcanzar valores estables de RMSD en tiempos menores a los de los modelos que incluyen a RsmZ, se consideró que sólo los primeros 100 ns constituyen la etapa de equilibración. En consecuencia, la etapa de producción para el modelo RsmE está constituida por 10 simulaciones de 100 ns en lugar de 50 ns. Así, los modelos RsmZ-(RsmE)$_3$, RsmZ-(RsmE)$_2$, RsmZ^{m1}-(RsmE)$_2$, y RsmZ^{m2}-(RsmE)$_2$ presentan *sets* con 2000 *snapshots*, mientras que el *set* de RsmE libre contiene 4000 *snapshots*.

V.2 Simulaciones de *Umbrella Sampling*

Los cálculos de *Umbrella Sampling* se emplearon para forzar la separación de RsmE del MU que contiene a SC$_{GGA}$. Este MU está formado por los residuos A$_{38}$-G$_{39}$-G$_{40}$-A$_{41}$-C$_{42}$ (ver Figura I.3). El punto de partida de estas simulaciones fue la estructura equilibrada del modelo RsmZ-(RsmE)$_3$. La coordenada de reacción χ fue definida como la distancia entre dos centros de masa: Uno de ellos corresponde a los heteroátomos de la región simple cadena de RsmZ antes mencionada. El otro está formado por los C$_\alpha$ de $\beta 1_A$, $\beta 4_B$ y $\beta 5_B$ de RsmE. La Figura V.3 muestra una representación pictórica de los átomos utilizados para definir χ.

El valor de χ se incrementó de 19.7 Å a 35.0 Å utilizando un espaciamiento de 0.1 Å entre cada ventana. La última estructura de cada simulación fue utilizada como la conformación inicial de la siguiente. La constante de fuerza del potencial armónico de sesgo fue de 300.0 kcal/molÅ^2 (Ver Ecuación III.8). Cada ventana fue simulada durante 8 ns. Sin embargo, los primeros 5 ns fueron considerados como un período de equilibración. En consecuencia, sólo los últimos 3 ns de cada ventana fueron utilizados para calcular el PEL del evento. Las estructuras de la trayectoria fueron guardadas cada 40 ps. El tiempo total de simulación, así, alcanza los 1.22 μs.

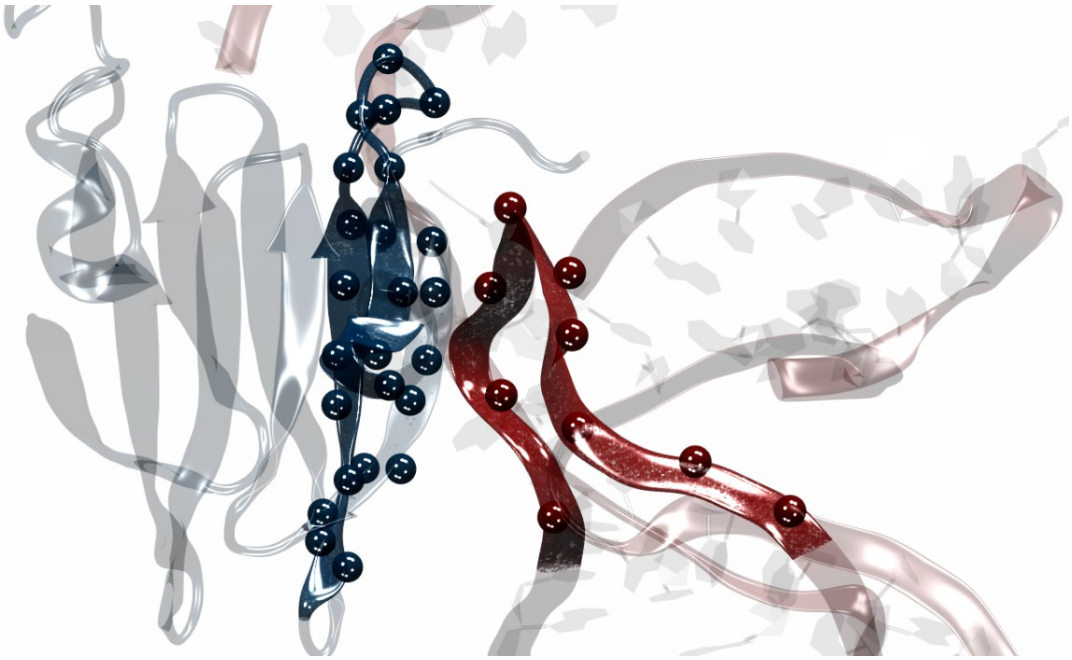


Figura V.3: Átomos utilizados para definir la coordenada de reacción χ como la distancia entre dos centros de masa. El correspondiente al SU de RsmE fue definido por los C_α representados como esferas azules. Por su parte, el correspondiente a RsmZ fue definido incluyendo los átomos P representados por esferas rojas. Los esqueletos que contienen a los átomos involucrados en la definición están resaltados en azul y rojo, respectivamente. Imagen adaptada de la Referencia 3.

Tres técnicas alternativas se utilizaron para calcular el PEL del evento a partir de los datos muestreados en las simulaciones sesgadas: WHAM [284], DHAM [285], y BAR [286] (Ver Sección III.4). En todos los casos, se emplearon códigos propios para remover el sesgo sobre las distribuciones obtenidas en los cálculos de *Umbrella Sampling*. Para WHAM y DHAM, la coordenada de reacción fue segmentada en fracciones de 0.025 Å. El algoritmo de WHAM requirió 200000 iteraciones para alcanzar una convergencia satisfactoria en el perfil, mientras que en el caso de DHAM las iteraciones se aplicaron sobre dos partes solapadas de la curva, ya que el programa falla por problemas de dimensión de matrices, al ser aplicado a todo el *set* de datos. Con el algoritmo de BAR, se emplearon 300 iteraciones para computar la diferencia de Energía Libre entre distribuciones adyacentes.

La Figura V.4 muestra el PEL para la separación entre RsmE y el motivo SC_{GGA} de RsmZ. En la esquina superior izquierda, se presenta una ilustración esquemática de las conformaciones de las moléculas interactuantes al comienzo del evento, mientras que en la esquina inferior derecha se muestran sus disposiciones al final del proceso de separación. La Energía Libre de unión estimada es de 35.4±2.0 kcal/mol. En este punto cabe recordar que las simulaciones de cada ventana de *Umbrella sampling* están equilibradas en todas las direcciones, excepto a lo largo de la coordenada de reacción. Por lo tanto, cuando la información para todas las ventanas se analiza en forma conjunta, se obtiene una descripción del proceso en cuasi-equilibrio. De esta manera, las simulaciones de *Umbrella Sampling* del proceso de separación entre dos fragmentos proveen información sobre el reconocimiento y la unión de los mismos cuando la información se analiza en sentido opuesto al simulado.

Ensayos de distinta naturaleza se realizaron para examinar la fiabilidad de los resultados. Los mismos se encuentran disponibles en el Anexo de este Capítulo. Las Figuras V.12 y V.15 de dicha sección muestran la consistencia de los resultados, dando soporte al PEL presentado en la Figura V.4 y a las observaciones realizadas acerca del proceso simulado.

V.3 Mecanismo de unión entre RsmE y RsmZ

La animación que muestra la unión de RsmE al MU de la región SC de RsmZ está disponible en el siguiente enlace: `https://www.youtube.com/watch?v=rzm6ZY_C_uI`. En ella se aprecia claramente que los residuos Lys38, Met1 y Leu2 de RsmE forman sus interacciones con los nucleótidos del MU de RsmZ al principio del proceso de unión. Por otra parte, la Tabla V.1 muestra el número de contactos de los nucleótidos del segmento $A_{38}-C_{42}$ de RsmZ con el SU de RsmE, tal y como se observan en el complejo RsmZ-$(RsmE)_3$. Estas interacciones fueron determinadas partiendo de las estructuras disponibles en el *Protein Data Bank* (códigos de acceso 2mf0 y 2mf1). Se observa que el residuo G_{40} es el que posee más interacciones. Éste forma seis puentes de Hidrógeno, seguido de G_{39} y A_{41} con tres interacciones cada uno. Cerrando la lista aparecen A_{38}, con dos puentes de Hidrógeno, y C_{42} con

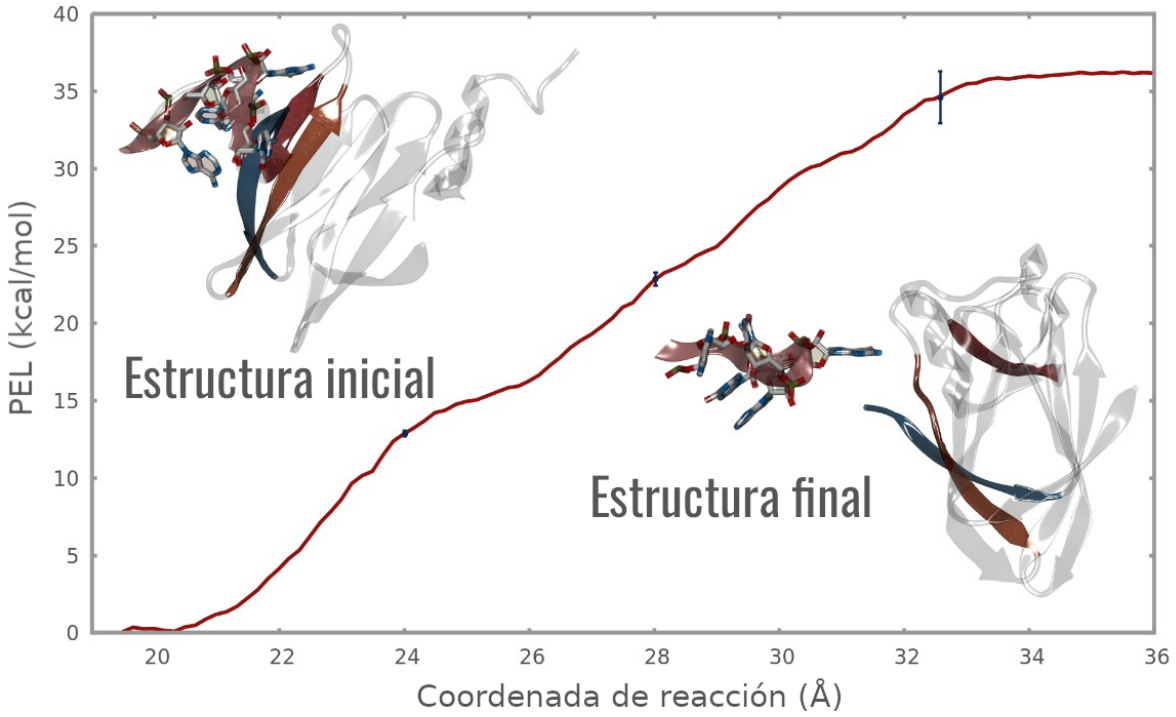


Figura V.4: Potencial de fuerza media para la separación entre RsmE y la secuencia $A_{38}G_{39}G_{40}A_{41}C_{42}$ presente en el MU de RsmZ. Una representación esquemática de la conformación inicial y final de las moléculas se encuentra también presente. Imagen adaptada de la Referencia 3.

sólo uno.

Habiendo identificado las interacciones más relevantes en cada extremo del rango de la coordenada de reacción, queda por dilucidar lo que ocurre en etapas intermedias. Con este fin, se hizo uso del módulo CPPTRAJ de AMBER para detectar los contactos entre el SU de RsmE y los nucleótidos de la secuencia $A_{38}-C_{42}$ de RsmZ. Se consideró que existe un contacto entre ambos grupos cuando la distancia entre cualquier átomo del SU de RsmE y cualquiera de los nucleótidos del MU de RsmZ es menor a 4.5 Å. La Figura V.5 muestra la evolución de este número de contactos a lo largo del proceso de separación/unión de ambas moléculas.

El nucleótido con mayor cantidad de contactos a lo largo del evento es G_{40},

Nucleótido	$\beta 1_B$	$\beta 5_A$	$\beta 4_A$	Loop $\beta 4_A$-$\beta 5_A$
A_{38}	Thr5	Arg44	-	-
G_{39}	-	Val42, His43, Arg44	-	-
G_{40}	Met1	Val40, Val42	Ala36,	Pro37 , Lys38
A_{41}	Met1, Ile3 (twice)	-	-	-
C_{42}	-	-	Arg31	-

Tabla V.1: Interacciones de puente de Hidrógeno entre el SU de RsmE y el pentanucleótido $A_{38}-C_{42}$ de RsmZ observadas en las estructuras cuyos códigos de acceso PDB son 2mf0 y 2mf1.

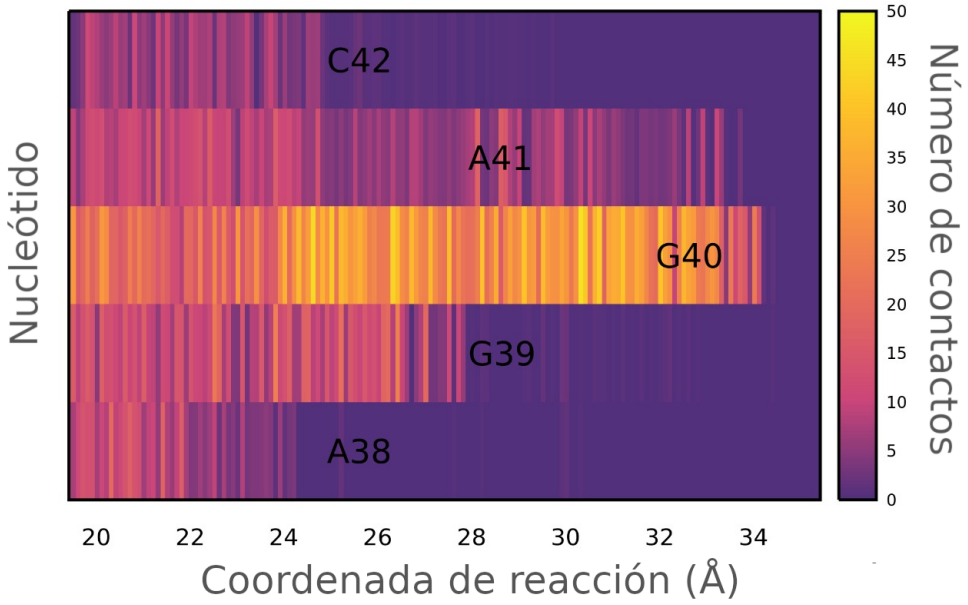


Figura V.5: Número de contactos entre los nucleótidos del motivo $A_{38}-C_{42}$ con el SU de RsmE. En la escala de colores, mientras más amarillo sea el color, mayor es el número de contactos. Los colores violeta, por su parte, indican una cantidad nula de contactos. Imagen adaptada de la Referencia 3.

el cual se mantiene próximo al SU de RsmE hasta muy avanzado el proceso de separación. Teniendo en cuenta la reversibilidad de las simulaciones de *Umbrella sampling*, esto indica que G_{40} participa de las primeras interacciones entre RsmE y RsmZ. Cabe destacar que la posición G_{40} en la secuencia AGGAC ya había sido señalada como una de las más importantes para el proceso de unión, en un estudio

realizado sobre moléculas pequeñas con capacidad de inhibir la acción de RsmE [80]. Los resultados de esta sección revelan que G_{40} tiene un rol fundamental en el reconocimiento de la proteína.

Asimismo, A_{41} está también en contacto con RsmE en un amplio rango de valores de χ, aunque presenta una cantidad significativamente más baja de contactos que G_{40}. Esto sugiere que A_{41} tiene también un rol importante en el reconocimiento entre ambas moléculas. Por el contrario, los residuos A_{38} y C_{42}, a cada lado del MU, presentan el menor número de interacciones de toda la secuencia, y las mismas se anulan para valores de $\chi > 25$ Å. Finalmente, G_{39} tiene un número de contactos significativamente mayor que A_{41}, pero los mismos desaparecen a valores menores de χ (≈ 28 Å). Esto demuestra que G_{39} estabiliza a las moléculas unidas, pero no participa de las etapas iniciales del proceso de unión. Al observar estos resultados, y compararlos con el análisis resumido en la Tabla V.1, una conclusión preliminar que emerge es que existe una clara correlación entre el número de puentes de Hidrógeno de los nucleótidos en el estado unido, y su involucramiento en el reconocimiento de RsmE.

La Figura V.6 completa el análisis de la Figura V.5 mostrando las interacciones de los residuos que componen el SU de RsmE con la secuencia $A_{38}-C_{42}$ de RsmZ. El panel superior muestra que el residuo Lys38, localizado en el *loop* entre $\beta 4_B$ y $\beta 5_B$, interactúa con RsmZ hasta el final del proceso de separación. Esto implica que se encuentra involucrado en los primeros pasos de reconocimiento entre ambas moléculas. Algo similar ocurre con los residuos Met1 y Leu2, localizados en el origen de $\beta 1_A$. Sin embargo, el número de contactos de Met1 disminuye a valores intermedios de χ, mientras que los correspondientes a Leu2 se mantienen elevados a lo largo de todo el rango de la coordenada. Los aminoácidos de $\beta 5_B$ no establecen tantos contactos como aquellos de $\beta 1_A$ y $\beta 4_B$, y estos no presentan contactos a grandes valores de χ. No obstante, asumen un rol más relevante a valores intermedios del proceso. En particular, el residuo Arg44 presenta varias interacciones en el estado unido, mientras Val42 y His43 forman más interacciones en etapas intermedias del evento.

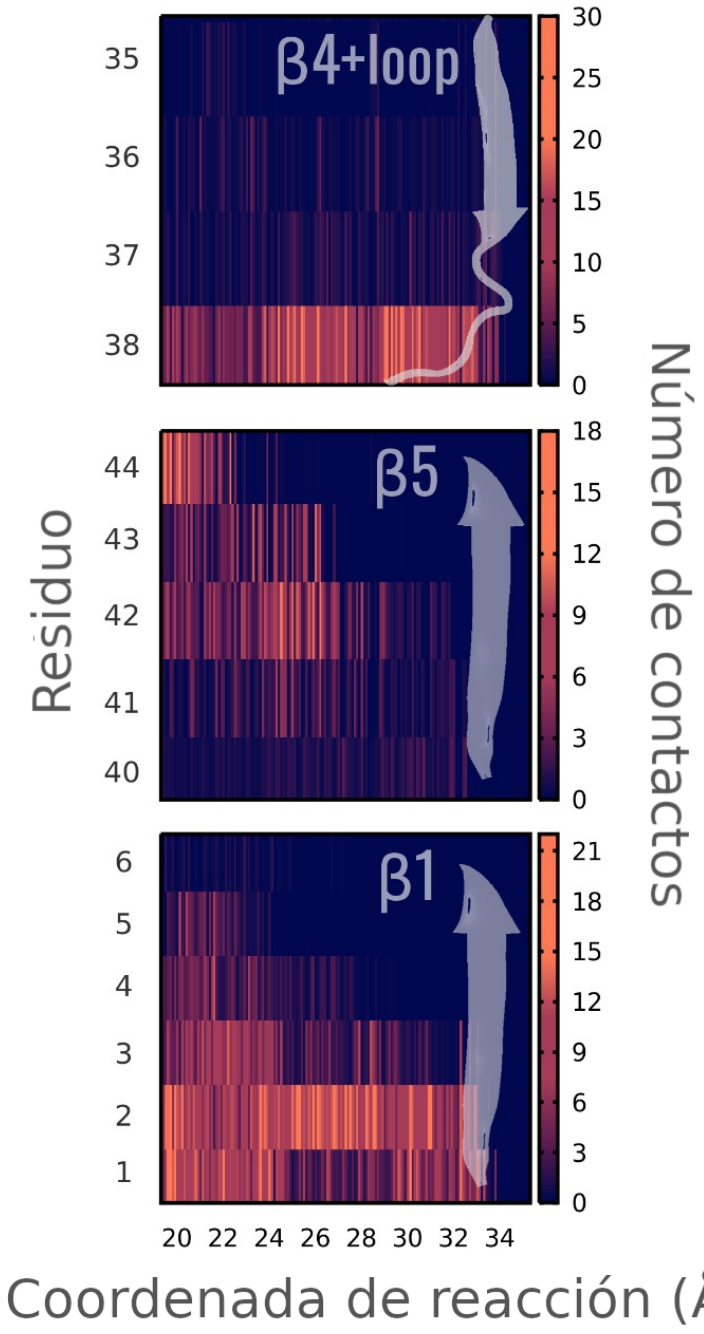


Figura V.6: Número de contactos para los principales residuos del SU de RsmE con el MU de RsmZ a lo largo del proceso de separación. Imagen adaptada de la Referencia 3.

V.4 Cambios conformacionales durante el proceso de unión

La película que representa la salida de RsmE de RsmZ-$(RsmE)_3$ muestra que ambas moléculas se deforman significativamente a lo largo del proceso. La deformación más importante ocurre en la región de simple cadena de RsmZ, que se desplaza acompañando los movimientos de RsmE. Asimismo, RsmE también presenta deformaciones en su estructura a lo largo del evento. En las siguientes secciones se describen las mediciones realizadas para cuantificar estos cambios conformacionales.

V.4.1 Deformaciones de RsmE

Durante las simulaciones de *Umbrella Sampling*, las láminas β_{1A} y β_{4A} de RsmE se alejan del centro de la proteína en las etapas intermedias, lo que hace que aumente el ángulo formado entre β_{1A} y β_{2A}, así como aquel formado por β_{4B} y β_{5B}. Para cuantificar estas observaciones se estimó que el ángulo entre $\beta 1_A$ y $\beta 2_A$ está descrito por aquel formado entre los C_α de Leu4, Val8 e Ile12. De forma similar, el ángulo entre $\beta 4_B$ y $\beta 5_B$ fue calculado considerando los C_α de Ile34, Lys38 y Arg44. La evolución de los ángulos formados entre $\beta 1_A$ y $\beta 2_A$, así como entre $\beta 4_B$ y $\beta 5_B$, se muestran en la Figura V.7. Estos valores son comparados con el mismo parámetro calculado sobre el SU libre de la proteína.

Se observa que las láminas $\beta 1_A$ y $\beta 4_B$ se separan respecto del resto de la proteína en etapas intermedias del proceso de separación, aumentando el ángulo entre ellas respecto a $\beta 2_A$ y $\beta 5_B$, respectivamente. Cuando RsmE se encuentra unido a RsmZ, el ángulo entre $\beta 1_A$ y $\beta 2_A$ fluctúa alrededor de $40° \pm 5°$. Conforme avanza el proceso de separación, éste valor aumenta hasta alcanzar un máximo cercano a $\chi \approx 29$ Å, donde oscila en torno a $48° \pm 5°$. A partir de este punto, en el rango $29 < \chi < 31$ Å, fluctúa y alcanza un nuevo máximo alrededor de $\chi \approx 33$ Å. Finalmente, decae a valores cercanos al original. La misma medición aplicada a la cadena contraria de la proteína no muestra variaciones significativas a lo largo del evento.

Por su parte, el ángulo entre $\beta 4_B$ y $\beta 5_B$ fluctúa en torno a los 45° al comienzo del proceso, para luego aumentar sostenidamente hasta alcanzar su máximo en

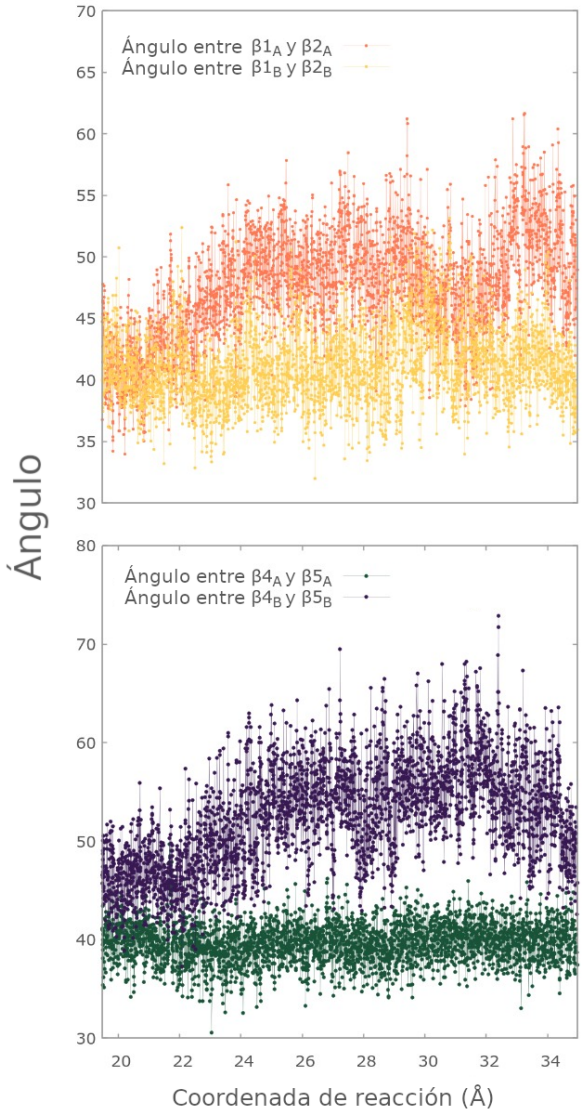


Figura V.7: Parámetros que describen la deformación de RsmE durante el proceso de salida del complejo RsmZ-(RsmE)$_3$. **Panel superior:** Variación del ángulo entre $\beta 1_A$ y $\beta 2_A$, y entre $\beta 1_B$ y $\beta 2_B$, que corresponde al SU libre de RsmE (amarillo). **Panel inferior:** Variación del ángulo entre $\beta 4_B$ y $\beta 5_B$ (azul), y entre $\beta 4_A$ y $\beta 5_A$, que corresponde al SU libre de la proteína (verde). Imagen adaptada de la Referencia 3.

valores de χ por encima de ≈31-32Å, donde oscila en torno a ≈57°. Para valores de χ aún mayores, este ángulo disminuye significativamente y termina oscilando alrededor de los 50° cuando los fragmentos ya están casi separados. De la misma manera que para el ángulo entre $\beta 1_A$ y $\beta 2_A$, la cadena contraria de la proteína no muestra variaciones significativas a lo largo del evento.

V.4.2 Deformaciones de RsmZ-(RsmE)$_2$

La mayor deformación de RsmZ que puede apreciarse a simple vista en la película es que la región que contiene al motivo SC$_{GGA}$ adopta una conformación similar a un SL durante el reconocimiento de RsmE. Esta observación concuerda con estudios experimentales previos que sugieren que este tipo de deformación es necesaria para que los segmentos simple cadena del ARN puedan ser detectados por RsmE [26]. Además, el fenómeno es coherente con la conformación que poseen los demás MU del sistema, y complementa los resultados referidos a la exposición del segmento SC$_{GGA}$ presentados en la Sección IV.3.3 del Capítulo IV. Precisamente, en ese capítulo se describe que la principal diferencia entre los modelos "L2R2-SC-accesible" y "L2R2-SC-ocluido" es la interacción entre los residuos C37 y A41 del segmento SC. Esta interacción, sólo presente en el modelo L2R2-SC-accesible, es una de las involucradas en la adopción de la forma similar a un SL.

Para cuantificar esta deformación observada en RsmZ, se midió la distancia entre dos centros de masa: uno de ellos está formado por los nucleótidos del motivo SC$_{GGA}$, mientras que el otro se constituye por los nucleótidos 16-18 y 56-58, localizados en las zonas que unen a SL1-SL2 y SL3-SL4, respectivamente. Estas últimas regiones definen, aproximadamente, el centro de RsmZ. Además, para evaluar la torsión del MU, se midió el ángulo formado por el centro de masa entre las bases de los residuos A$_{38}$-G$_{40}$-C$_{42}$. La Figura V.8 muestra la evolución de ambos parámetros durante el proceso.

La distancia entre el motivo SC$_{GGA}$ y el centro de RsmZ oscila en ≈12-14Å cuando las moléculas están fuertemente unidas entre sí. Posteriormente, presenta un aumento significativo mientras la coordenada de reacción varía entre 22Å y 25Å, para luego oscilar alrededor de ≈23Å hasta valores de χ de ≈32Å. Al final del

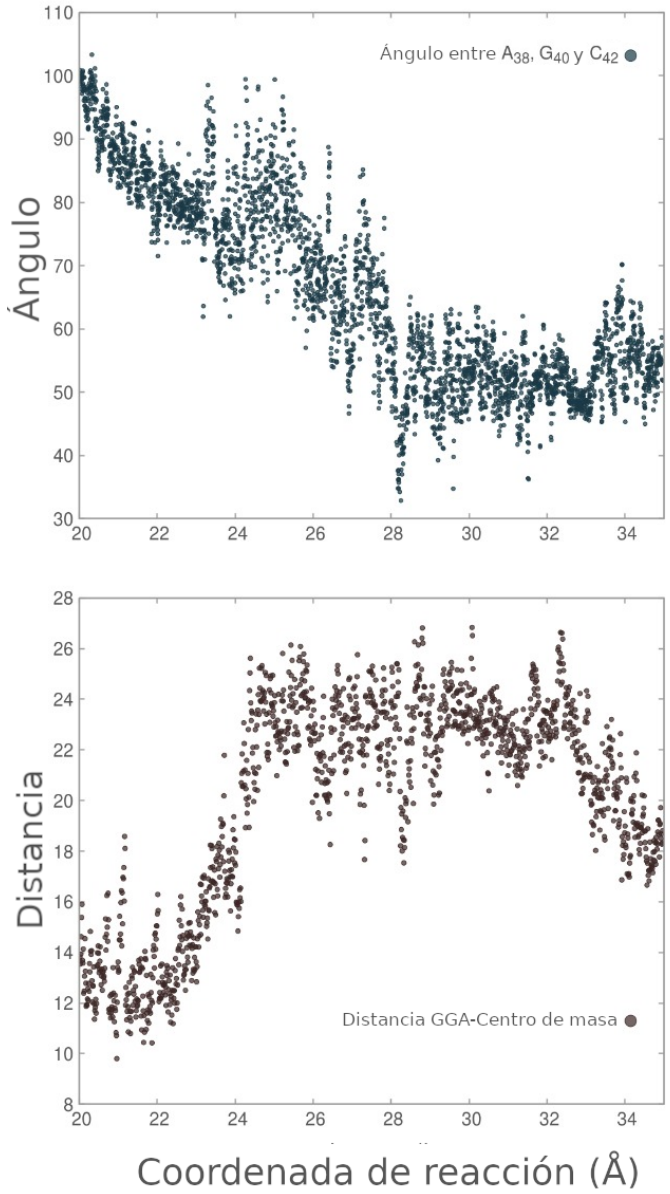


Figura V.8: Parámetros que describen las deformaciones de RsmZ a medida que se separa de RsmE. **Panel Superior:** Variación del ángulo entre los nucleótidos A_{38}, G_{40} y C_{42}. **Panel Inferior:** Variación de la distancia entre el centro de masa del motivo SC_{GGA} y el centro de RsmZ. Imagen adaptada de la Referencia 3.

proceso, cuando RsmE está prácticamente separado de RsmZ, este valor decrece nuevamente a ≈18Å. Por su parte, el ángulo entre los nucleótidos A_{38}-G_{40}-C_{42} decrece de manera sostenida, desde que comienza el proceso de separación hasta que χ alcanza un valor de ≈ 28Å, cuando oscila entre 30-40°. En conjunto, estas observaciones describen cómo el *loop* se cierra a medida que el evento ocurre, adoptando la forma similar a la de un SL descrita párrafos atrás.

V.5 Deformaciones de RsmE y RsmZ-$(RsmE)_2$ aisladas

Una conclusión preliminar que emerge de los resultados presentados en las secciones anteriores es que las deformaciones sufridas por RsmE y RsmZ-$(RsmE)_2$ son necesarias para las primeras etapas del proceso de unión. Por lo tanto, resulta pertinente estudiar si cada molécula, por separado, puede adoptar conformaciones similares a las requeridas al momento de iniciarse el reconocimiento.

Para analizar estas características estructurales, se midió la distancia entre el motivo SC_{GGA} y el centro de RsmZ, así como también el ángulo A_{38}-G_{40}-C_{42} en las estructuras muestreadas a lo largo de las 10 trayectorias de RsmZ-$(RsmE)_2$. Los mismos parámetros fueron también medidos en los modelos $RsmZ^{m1}$-$(RsmE)_2$ y $RsmZ^{m2}$-$(RsmE)_2$, con el objetivo de poner a prueba cómo influye la longitud de esta región en su capacidad para adoptar las conformaciones observadas durante la unión. Por otro lado, se evaluaron los ángulos entre $\beta 1_A$ y $\beta 2_A$ y el que forman $\beta 4_B$ y $\beta 5_B$ en las estructuras obtenidas en las 10 simulaciones de RsmE. Dichos valores fueron utilizados para estimar las densidades de probabilidad de las distancias y ángulos considerados.

Las probabilidades para la distancia entre el motivo SC_{GGA} y el centro de RsmZ en la estructura nativa y en las mutantes se muestran en el panel superior de la Figura V.9. Las distribuciones para el ángulo A_{38}-G_{40}-C_{42} están presentes en el panel inferior. La Figura V.10 muestra la superposición de ese segmento en los tres modelos. Dicha figura fue construida con las estructuras que presentan la máxima distancia entre el motivo GGA y el centro de masa de RsmZ de cada modelo.

La Figura V.9 muestra que la distribución correspondiente a la distancia entre el motivo SC_{GGA} y el centro de RsmZ-$(RsmE)_2$ es multimodal y significativamente

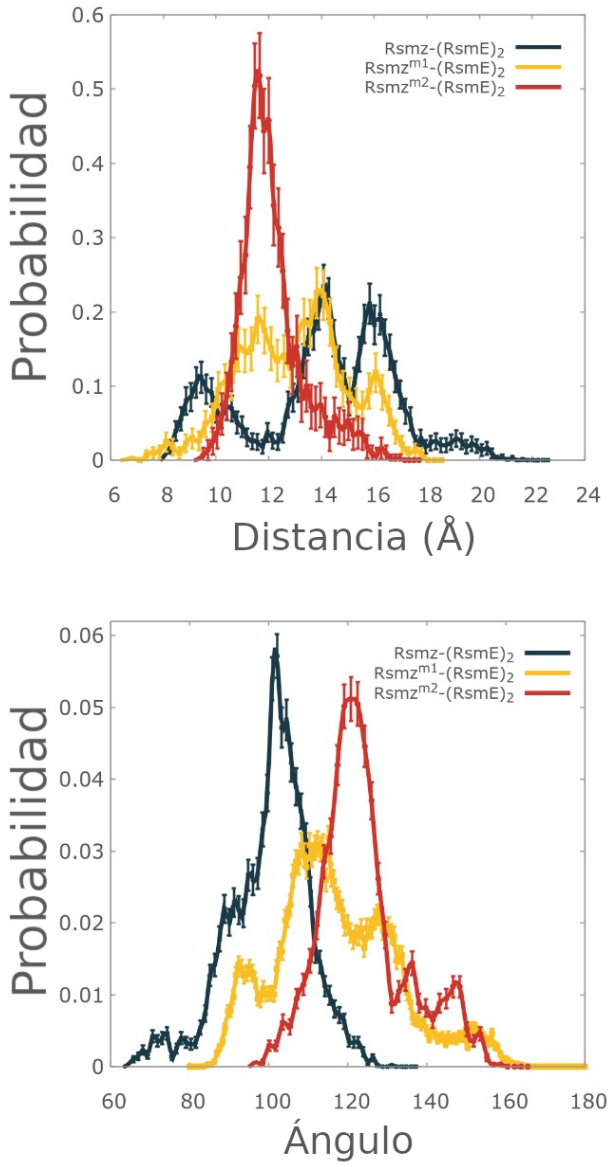


Figura V.9: **Panel superior:** FDP para la distancia entre el motivo SC$_{GGA}$ y el centro de RsmZ en los modelos estudiados en este capítulo. **Panel inferior:** FDP para el ángulo A$_{38}$-G$_{40}$-C$_{42}$ medido en Rsmz-(RsmE)$_2$ (azul), Rsmz^{m1}-(RsmE)$_2$ (amarillo) y Rsmz^{m2}-(RsmE)$_2$ (rojo). Los datos corresponden a las 10 simulaciones de DM para el complejo Rsmz-(RsmE)$_2$ y sus mutantes. Imagen adaptada de la Referencia 3.

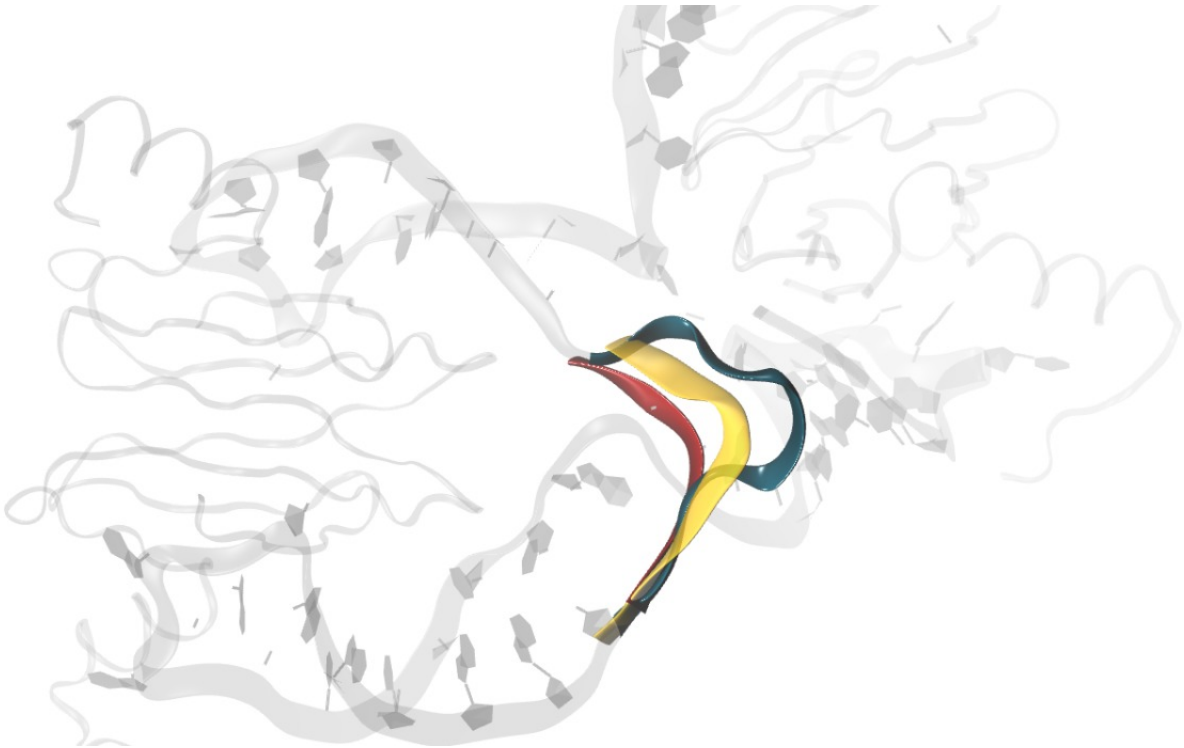


Figura V.10: Superposición de la región SC de los modelos estudiados en este capítulo, utilizando las estructuras que presentan el máximo valor para la distancia entre el motivo SC_{GGA} y el centro de masa de RsmZ de sus respectivas simulaciones. Azul: RsmZ-$(RsmE)_2$, Amarillo: $RsmZ^{m1}$-$(RsmE)_2$, Rojo: $RsmZ^{m2}$-$(RsmE)_2$. Imagen adaptada de la Referencia 3.

amplia. Presenta un pico a ≈8Å, otro alrededor de ≈14Å, y un tercer pico en ≈16Å. Pasado ese punto la probabilidad decae, aunque mantiene valores pequeños pero significativos hasta alrededor de 21 Å. Esta distribución da cuenta de una gran flexibilidad para este movimiento. Sin embargo, la distancia nunca alcanza los valores observados al inicio del proceso de unión (≈26 Å). De manera similar, el panel inferior de la figura muestra que el MU del complejo RsmZ-$(RsmE)_2$ libre no llega a flexionarse tanto como se requiere al inicio del proceso de unión (≈45°). Sin embargo, las estructuras que contribuyen a la cola de la distribución se acercan a dichos valores. Así, el MU requiere sólo una mínima contribución por parte de RsmE para lograr exponer a G_{40}.

Las distribuciones correspondientes a los modelos mutados son marcadamente diferentes. En las simulaciones de Rsmz^{m1}-(RsmE)$_2$, el ángulo A$_{38}$-G$_{40}$-C$_{42}$ nunca alcanza valores menores a 80°, mientras que para Rsmz^{m2}-(RsmE)$_2$ infrecuentemente llega a valores menores a 100°. Algo similar ocurre con las distribuciones de distancias entre el motivo SC$_{GGA}$ y el centro de masa de RsmZ. Para el caso de Rsmz^{m1}-(RsmE)$_2$ existe un pico a ≈12 Å que no se encuentra presente en el modelo silvestre, al mismo tiempo que el pico alrededor de ≈16 Å se vuelve más infrecuente. Asimismo, el modelo Rsmz^{m2}-(RsmE)$_2$ presenta un máximo precisamente a ≈12 Å, mientras que los picos a valores mayores desaparecen. Tanto la imposibilidad de llegar a conformaciones con ángulos A$_{38}$-G$_{40}$-C$_{42}$ menores a 80°como con distancias SC$_{GGA}$-centro de masa de RsmZ mayores a 18 Å , indican que los modelos mutados son menos flexibles que la forma silvestre, sugiriendo que la región de simple cadena de RsmZ requiere una extensión mínima de 9 nucleótidos para poder deformarse y exponer al nucleótido G$_{40}$.

La Referencia 73 señala que la región de simple cadena entre SL2 y SL3 en la forma nativa de RsmZ no tiene la longitud óptima para maximizar la afinidad del bolsillo de unión a RsmE formado por dichos *stem-loops*. En particular, el estudio muestra que una mutación de RsmZ en la que el empalme entre SL2 y SL3 se acorta en tres nucleótidos resulta en una afinidad tres veces mayor por el primer dímero de RsmE, y una mayor activación de la traducción de los genes regulados por ésta. La hipótesis surgida de ese trabajo es que el empalme entre SL2 y SL3 evolucionó de manera de contener un motivo de unión adicional en lugar de maximizar la afinidad de unión. Los resultados presentados en este capítulo sugieren que esos tres nucleótidos adicionales, que provocan una funcionalidad subóptima de RsmZ para capturar a la primera unidad de RsmE, son requeridos para conferir al ARNp la habilidad de capturar a la tercera unidad de la proteína.

V.5.1 Análisis de Componentes Principales

Las macromoléculas suelen realizar movimientos concertados de gran amplitud. Estos se encuentran dentro de un subespacio de relativamente pocas dimensiones, caracterizado por tener una gran flexibilidad [288]. Este subespacio se denomina como

el EE del sistema, y contiene los movimientos necesarios para su funcionamiento [304]. La metodología de PCA permite determinar el EE de una macromolécula mediante la diagonalización de la matriz de covarianza calculada a partir de las conformaciones surgidas de simulaciones de DM.

En la sección anterior se describió que, cuando RsmZ-$(RsmE)_2$ y RsmE se encuentran aisladas, realizan movimientos que tienden a generar las conformaciones requeridas para el proceso de interacción. Sin embargo, en esas simulaciones, no se alcanzaron deformaciones equivalentes a las observadas durante las etapas iniciales del proceso de unión. Para estudiar si las direcciones de los movimientos detectadas en la etapa de reconocimiento entre RsmZ-$(RsmE)_2$ y RsmE se encuentran contenidas en el EE de estas moléculas, se realizó un estudio de PCA para cada una de ellas.

En el caso de RsmE, la matriz de covarianza fue computada utilizando las coordenadas cartesianas de los átomos C_α correspondientes a los residuos 1 a 44 de ambas cadenas. Esta selección excluye a las α-hélices de los extremos C-terminal. Estos segmentos tienen una alta flexibilidad, pero no están involucrados en el proceso de unión, de manera que su inclusión en el análisis perjudica la identificación de las fluctuaciones relevantes para el evento estudiado [303]. Para el complejo RsmZ-$(RsmE)_2$, la matriz fue calculada considerando las coordenadas cartesianas de los átomos de Fósforo de RsmZ. La utilización de estas coordenadas cartesianas en lugar de los ángulos diedros utilizados en el Capítulo IV responde al hecho de que la presencia de dos moléculas de RsmE ya unidas a RsmZ le confieren suficiente rigidez como para permitir dicho análisis. Al igual que en el Capítulo IV, el análisis se realizó sobre la trayectoria concatenada de 10 simulaciones independientes del sistema para mejorar la consistencia de los resultados [302].

La animación del primer componente principal de RsmZ está disponible en el siguiente enlace: https://www.youtube.com/watch?v=ufBU9Jsqpug. Este componente, o modo colectivo, describe un movimiento tipo *bending* cuyo vértice es el centro de la molécula y sus extremos son los *stem-loops* SL2/SL3 por un lado y los correspondientes a SL1/SL4, por otro lado. La ubicación de SL1 relativa a SL4, así como la correspondiente a SL2 en relación a SL3, permanecen casi

inalteradas ya que cada par de SLs se encuentra ocupado por RsmE (aunque las dos moléculas de RsmE no están representadas en la película para poder observar de forma más clara el movimiento de RsmZ). En el siguiente enlace se muestran los desplazamientos correspondientes al segundo componente principal: https://www.youtube.com/watch?v=oDlmLTYYRu4. Como puede verse, estos modifican la posición del MU en la región SC que une SL2 y SL3, la cual oscila entre una posición expuesta y una oculta. Así, el modo muestra cómo alternativamente se expone y oculta el MU SC$_{GGA}$.

Finalmente, la animación para el primer modo de RsmE está disponible en el enlace https://www.youtube.com/watch?v=jaOuTVPTKcc. Los átomos C$_\alpha$ de los residuos Leu4, Val8 e Ile12, que proporcionan una medida del ángulo formado entre $\beta 1_A$ y $\beta 2_A$, están representados como esferas anaranjadas. Por su parte, los C$_\alpha$ de los residuos Ile34, Lys38 y Arg44, usados para estimar el ángulo entre $\beta 4_B$ y $\beta 5_B$, aparecen coloreados en azul. Es evidente que esta coordenada colectiva consiste en una oscilación concertada de ambos ángulos.

V.5.2 Movimientos esenciales de RsmZ-(RsmE)$_2$ y RsmE

El análisis descrito en esta sección se empleó para evaluar en qué medida los movimientos colectivos del EE de RsmE y de RsmZ-(RsmE)$_2$ son capaces de generar las estructuras observadas en las etapas iniciales del proceso de unión, cuando se produce el reconocimiento entre ambos fragmentos y los mismos tienen las mayores deformaciones. Para realizarlo, se tomaron las estructuras de ambos fragmentos muestreadas para valores de χ entre 27.0 y 36.0 Å de las simulaciones de *Umbrella Sampling*. Este conjunto estructuras serán denominadas $\mathbf{X}^i_{tar}$ (donde el superíndice identifica a cada estructura del *set*, mientras que el subíndice *tar* viene del inglés *target*, ya que estas estructuras son el objetivo a alcanzar en el ajuste que se discute a continuación). Cabe destacar que hay un *set* de estructuras $\mathbf{X}^i_{tar}$ para RsmE y otro para el complejo RsmZ-(RsmE)$_2$. El proceso descrito debajo se aplicó separadamente a cada conjunto.

Primero, cada $\mathbf{X}^i_{tar}$ fue alineada a la estructura de referencia empleada para hacer el PCA correspondiente. Luego, para cada una de estas estructuras alineadas, se

buscaron los coeficientes para calcular $\mathbf{X}^i_{fit}$ de acuerdo con la Ecuación V.1. $\mathbf{X}^i_{fit}$ es la estructura más cercana a $\mathbf{X}^i_{tar}$, que se puede obtener como combinación lineal de los vectores del EE ($\bar{\mathbf{v}}_n$) partiendo del mismo origen, $\mathbf{X}_0$, que es la estructura promediada de la trayectoria empleada para determinar los PCA. Así:

$$\mathbf{X}^i_{fit} = \mathbf{X}_0 + \sum_{n=1}^{N_e} a^i_n \mathbf{v}_n. \tag{V.1}$$

Aquí n identifica los autovectores del PCA y N_e es la cantidad de autovectores incluidos en el EE. Los coeficientes de esta expansión, a^i_n, fueron elegidos para minimizar el valor de RMSD$_i$ entre $\mathbf{X}^i_{tar}$ y $\mathbf{X}^i_{fit}$. Cuanto menores sean los valores de RMSD$_i$, mayor es la capacidad del EE de las moléculas aisladas, de generar las conformaciones observadas al momento de iniciarse el reconocimiento entre ambas [305].

Las bondades de los ajustes se presentan en la Figura V.11, en donde se compara el RMSD$_i$ entre $\mathbf{X}^i_{tar}$ y $\mathbf{X}^i_{fit}$, con el correspondiente a $\mathbf{X}^i_{tar}$ y $\mathbf{X}_0$ (siendo éste el RMSD previo al ajuste). Para los cálculos, se consideraron los primeros 6 autovectores de RsmE y los primeros 4 para RsmZ-(RsmE)$_2$. Este número de vectores presenta, aproximadamente, un 3% del total de los grados de libertad de cada molécula, pero dan cuenta de ≈90% del total de las fluctuaciones observadas en la trayectoria concatenada. Por último, cabe mencionar que para RsmE, el ajuste no se realizó sobre toda la estructura, sino sólo sobre uno de sus SU. Esto se debe a que los componentes principales de RsmE libre son (casi) simétricos con respecto al plano de simetría del dímero. Sin embargo, las estructuras obtenidas de las simulaciones de *Umbrella Sampling* son perturbadas por un sólo lado.

Con fines comparativos, el mismo ajuste se realizó con conjuntos alternativos de vectores. Así, para RsmE, se utilizaron los vectores 7-12 y 20-25 del PCA. Cabe notar que los vectores 7-12 deberían proporcionar el segundo mejor ajuste, justo después del obtenido con los vectores 1-6. Siguiendo criterios similares, para RsmZ, se emplearon los vectores 5-9 y 20-23. La Figura V.11 muestra que, para cualquiera de los conjuntos alternativos, los valores de RMSD antes y después del ajuste son casi los mismos. Este resultado ilustra la falta de capacidad de estos

subespacios para generar las estructuras observadas al principio del evento de unión. En cambio, los EE de ambas moléculas proporcionan un ajuste considerablemente mejor, reduciendo el RMSD inicial en más de un 50%.

La Figura V.11 también ilustra la diferencia en la flexibilidad de cada molécula, siendo RsmE mucho más rígida que RsmZ. Este es un hallazgo típico al comparar la flexibilidad de los ARN con la de las proteínas globulares [295]. En este caso particular, la flexibilidad observada para RsmZ se reduce significativamente en comparación con la del ARNp libre, ya que se encuentra unido a dos dímeros de RsmE. Sin embargo, es mucho más flexible que RsmE. Una evaluación cuantitativa de estas diferencias se puede encontrar en las Tablas V.2 y V.3, las cuales presentan los autovalores de los componentes principales para ambos sistemas. Se observa que RsmZ, aún unido a dos moléculas de RsmE, posee valores marcadamente superiores a los de RsmE libre. Esto indica que sus desplazamientos son mayores que los de RsmE.

Modo	Autovalor (Å^2)	Porcentaje acumulado
1	1129.2	59.7
2	255.3	73.1
3	169.5	82.2
4	102.3	**87.5**

Tabla V.2: Autovalores correspondientes a los primeros cuatro PC del complejo RsmZ-(RsmE)$_2$ y sus porcentajes acumulados del desplazamiento total. Estos modos representan el 3.0% del total de los grados de libertad del sistema.

V.6 Conclusiones del capítulo

En este capítulo se presenta una descripción detallada del mecanismo de unión entre RsmE y el motivo AGGAC localizado en la región de simple cadena de RsmZ que une a SL2 y SL3. La misma estuvo basada en simulaciones de *Umbrella Sampling* que permitieron estudiar el proceso de unión/salida de RsmE del complejo RsmZ-(RsmE)$_2$. Los resultados muestran que G_{40}, el nucleótido central de la secuencia,

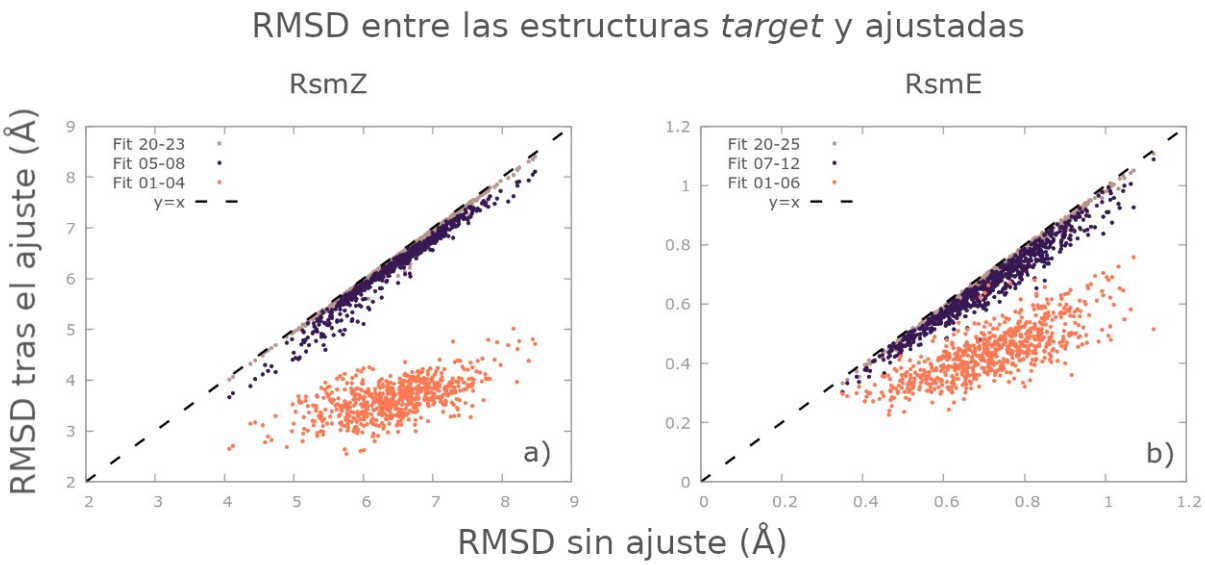


Figura V.11: RMSD entre $\mathbf{X}^i_{tar}$ y $\bar{\mathbf{X}}^i_{fit}$, comparado respecto al calculado entre $\mathbf{X}^i_{tar}$ y $\mathbf{X}_0$. $\mathbf{X}^i_{fit}$ es el mejor ajuste para $\mathbf{X}^i_{tar}$ deformando $\mathbf{X}_0$ en las direcciones de los vectores del EE. El panel A muestra los resultados para el SU de RsmE, mientras el panel B hace lo propio para RsmZ. Los puntos señalados en naranja corresponden al ajuste realizado a partir de los vectores empleados para la construcción del EE del sistema. Imagen adaptada de la Referencia 3.

Modo	Autovalor (Å^2)	Porcentaje acumulado
1	5.98	36.0
2	4.07	60.6
3	1.58	70.2
4	1.33	78.2
5	0.94	83.9
6	0.84	**89.0**

Tabla V.3: Autovalores correspondientes a los primeros seis componentes principales de RsmE y sus porcentajes acumulados del desplazamiento total. Estos modos representan el 3.0% del total de los grados de libertad del sistema.

está involucrado en las etapas iniciales del proceso, cuando las moléculas se reconocen entre sí. Además, este nucleótido establece el mayor número de contactos con RsmE a lo largo de todo el proceso. Por parte de RsmE, los aminoácidos que

participan en el reconocimiento son Met1 y Leu2 de $\beta 1_A$, y Lys38 de la interfaz entre $\beta 4_B$ y $\beta 5_B$. Los residuos localizados en $\beta 5_B$, sólo contactan a RsmE en estadíos intermedios o finales del proceso de unión.

Las simulaciones de DM estándar fueron utilizadas para caracterizar dinámicamente a RsmE y RsmZ-$(RsmE)_2$. Los resultados muestran que ambas moléculas por separado no alcanzan las conformaciones observadas al principio del proceso de unión. Sin embargo, los estudios de PCA sugieren que RsmE y RsmZ-$(RsmE)_2$ son fácilmente deformables a lo largo de coordenadas colectivas que tienden a generar las disposiciones requeridas durante la interacción. En el caso de RsmE, estos movimientos consisten mayormente en la separación entre $\beta 1$ y $\beta 4$ con respecto al resto de la proteína. Por su parte, para el complejo RsmZ-$(RsmE)_2$, hay dos coordenadas colectivas importantes: Una de ellas es un movimiento tipo *bending* del motivo AGGAC que ejecuta una oscilación del tipo "entrada-salida". La otra fluctuación importante, que es la segunda en cuanto a su amplitud, hace variar el ángulo formado entre los residuos A_{38}-G_{40}-C_{42}, permitiendo que la región adopte una conformación similar a un SL. Este movimiento permite que G_{40} sobresalga más allá de los demás nucleótidos del segmento, quedando en condiciones de reconocer más fácilmente a RsmE. Un comportamiento similar fue también descrito para los motivos AGGAX y AXGGAX presentes en otros ARN objetivo de RsmE [26]. Este resultado se condice con observaciones planteadas para los modelos L2R2-SC-accesible y L2R2-SC-ocluido en el Capítulo IV respecto a la alternancia entre una disposición expuesta y ocluida para dicha región el ARNp.

Las simulaciones de MD estándar fueron también empleadas para evaluar aspectos dinámicos de la región simple cadena de RsmZ entre SL2 y SL3 en los mutantes $RsmZ^{m1}$-$(RsmE)_2$ y $RsmZ^{m2}$-$(RsmE)_2$. Estos poseen dos y cuatro nucleótidos menos que la versión nativa de RsmZ, respectivamente. Estas simulaciones revelaron que los segmentos acortados no son lo suficientemente flexibles como para adoptar las conformaciones que el motivo AGGAC requiere al inicio del proceso de unión. Esto podría explicar por qué esta región es más larga que la que minimiza la constante de disociación de la primera unidad de RsmE (la cual se une al bolsillo formado por SL2 y SL3) [73].

En relación a las hipótesis planteadas al principio del capítulo, los resultados permiten confirmar ambos postulados: Por un lado, se verifica que la interacción entre RsmE y RsmZ-$(RsmE)_2$ es necesaria para que el ARNp adopte las conformaciones observadas al inicio del evento de unión, sugiriendo que es necesario cierto grado de ajuste inducido durante la unión. Por otra parte, y en relación a la segunda hipótesis, se comprobó que nucleótidos y residuos que ya habían sido sugeridos como críticos para la interacción entre RsmE y motivos A(X)GGAX, tienen un rol relevante durante todo el proceso de unión. Así, y habiendo cumplido con los objetivos planteados al principio del capítulo, se concluye que los resultados fueron satisfactorios en esta etapa de la investigación.

V.7 Anexo

Se realizaron diversos ensayos para estudiar la fiabilidad de los resultados obtenidos por *Umbrella Sampling*. En el primero de ellos se dividió el *set* completo de datos en tercios de igual tamaño, para luego calcular el PEL sobre cada uno de ellos. Posteriormente se calculó el desvío estándar entre los tres PEL en diversos puntos de la coordenada de reacción, para estimar la incerteza estadística de la curva obtenida. La Figura V.12 presenta los resultados de este ensayo.

También se compararon los perfiles obtenidos con WHAM, DHAM y BAR, observando que son prácticamente idénticos. Los resultados de esa comparación se presentan en la Figura V.13.

En otro de los ensayos, se puso a prueba la consistencia entre el PEL obtenido y las distribuciones de probabilidad sesgadas de cada ventana a partir de la cual ese PEL fue calculado. El ensayo compara las probabilidades sesgadas $P_i(\chi)$ directamente obtenidas por *Umbrella Sampling* en cada ventana i, con las probabilidades $\mu_i(\chi)$ estimadas a partir de la distribución desesgada obtenida con WHAM, añadiendo el efecto del potencial sesgado. La consistencia entre las dos distribuciones en cada ventana es evaluada por el Ensayo de Divergencia Simétrica de de Kullback-Leibler,

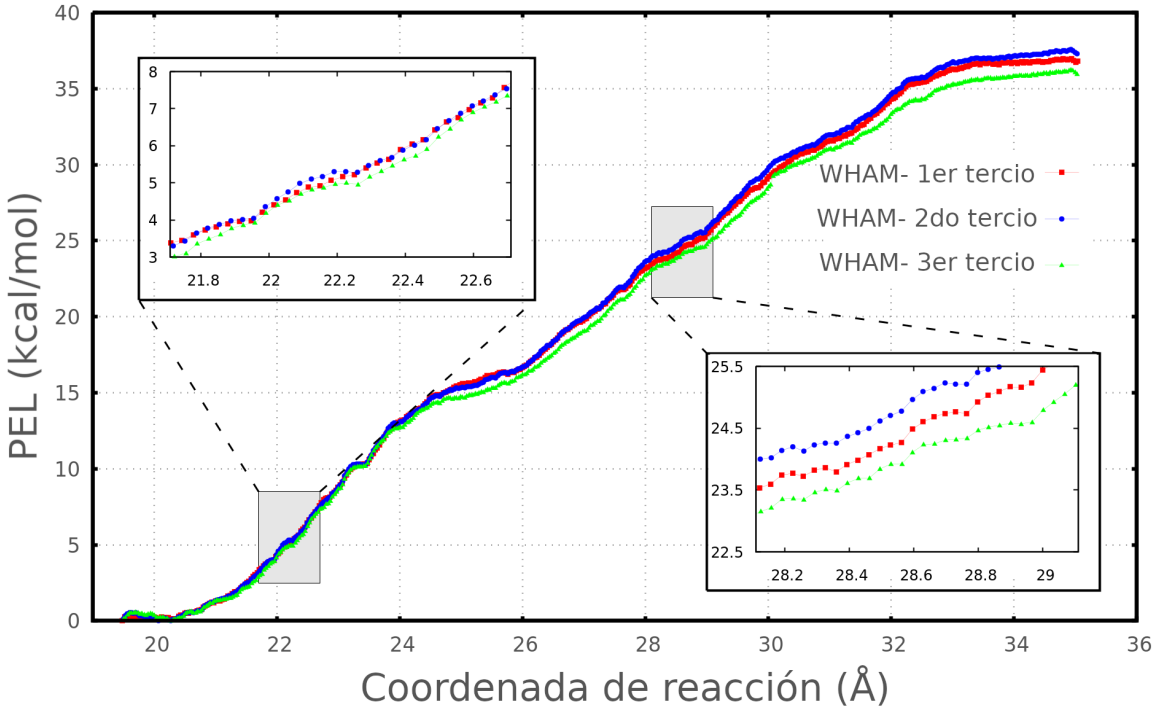


Figura V.12: Comparación de los perfiles obtenidos partiendo de tres fragmentos alternativos del *set* de datos, el cual fue dividido en partes iguales. Los recuadros muestran una amplificación de las tres curvas en dos segmentos ubicados en distintas partes del PEL para destacar la similaridad entre los valores de cada subconjunto. Imagen adaptada de la Referencia 3.

$$S_i = \frac{1}{2}\, D(P_i, \mu_i) + \frac{1}{2}\, D(\mu_i, P_i)\,, \tag{V.2}$$

donde,

$$D(f,g) = \sum_{k=1}^{N} f(\chi_k)\; ln\left(\frac{f(\chi_k)}{g(\chi_k)}\right). \tag{V.3}$$

En la última ecuación, N representa el número de segmentos empleados en la representación discretizada de la densidades de probabilidad $f(\chi)$ y $g(\chi)$, mientras χ_k es el valor de la variable aleatoria en el centro del segmento k. Cuanto menor sea

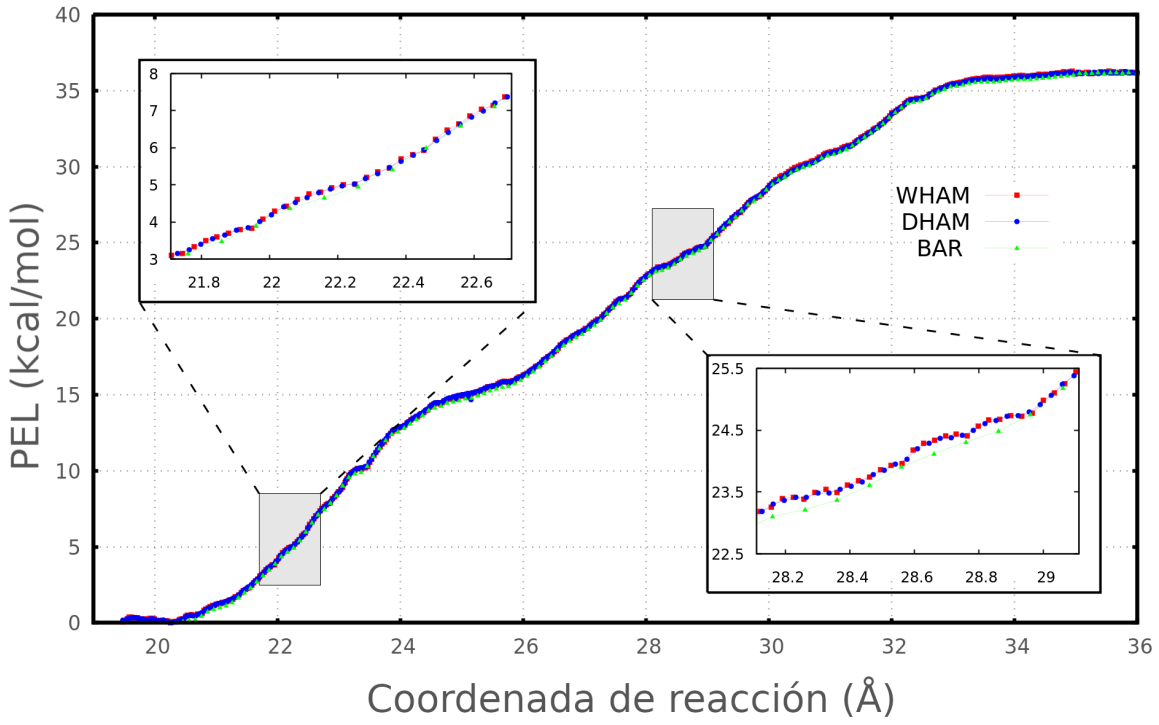


Figura V.13: Comparación de los PEL calculados con WHAM (rojo), DHAM (azul) y BAR (verde). Se muestran dos recuadros que amplifican segmentos del PEL en distintas partes del perfil para destacar la similaridad entre los valores obtenidos por los tres algoritmos. Imagen adaptada de la Referencia 3.

el valor de S_i, mejor es la consistencia entre las dos distribuciones. La Figura V.14 ilustra la Divergencia Simétrica de KL (Eq. V.2) como función de la coordenada de reacción para la salida de RsmE del MU situado en la región SC de RsmZ.

Finalmente, se utilizó la información recolectada en todos los pares de ventanas adyacentes para calcular:

$$f_i(\Delta U) = \ln\left(\frac{P_{i+1}(\Delta U)}{P_i(\Delta U)}\right) + \frac{\Delta U}{k_B T}. \tag{V.4}$$

Aquí, ΔU mide la diferencia entre los potenciales sesgados de las ventanas adyacentes, U_i y U_{i+1}, a un dado valor de χ situado entre los centros de ambas ventanas:

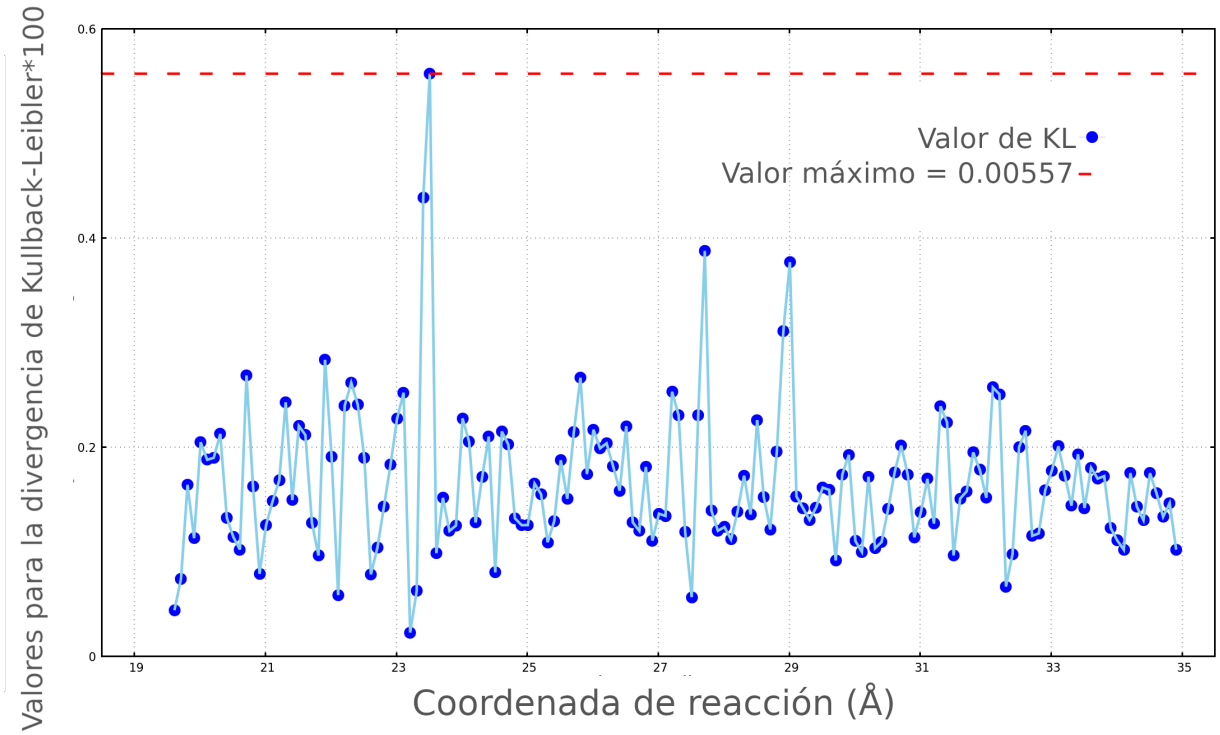


Figura V.14: Divergencia simétrica de Kullback-Leibler (S_i), definida en la Ecuación V.2 como función de la coordenada de reacción. Todos los valores de (S_i) están por debajo de 0.0058, indicando una consistencia razonable entre los datos obtenidos mediante *Umbrella sampling* con respecto a los arrojados por WHAM. Imagen adaptada de la Referencia 3.

$$\Delta U = \Delta U(\chi) = U_{i+1}(\chi) - U_i(\chi), \tag{V.5}$$

donde $P_i(\Delta U)$ y $P_{i+1}(\Delta U)$ son las distribuciones de probabilidad para χ obtenidas de las simulaciones i e $i+1$, respectivamente, pero expresadas como funciones de ΔU. Este cambio de variables puede realizarse debido a que una vez que los centros de distribuciones adyacentes fueron definidos, hay una correspondencia de uno a uno entre ΔU y χ. Por su parte, k_B es la constante de Boltzmann, y T es la temperatura absoluta utilizada en las simulaciones, 303 K. La función $f(\Delta U)$ debe ser constante en la región donde las distribuciones obtenidas para las ventanas i e $i+1$ se solapan [306]. Por el contrario, si la función tiene una pendiente definida,

eso indica que las muestras de ambas ventanas no arrojan resultados consistentes entre sí. La Figura V.15 muestra representaciones típicas para la función $f_i(\Delta U)$ (Eq. V.4).

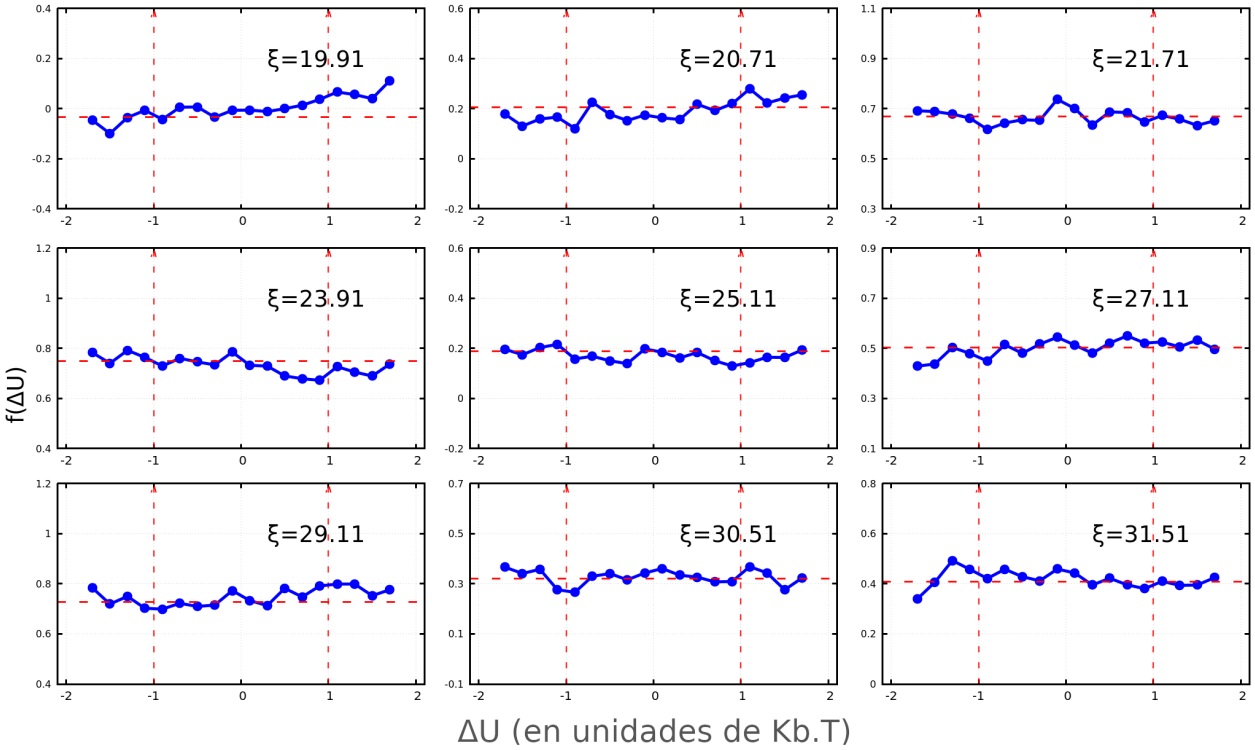
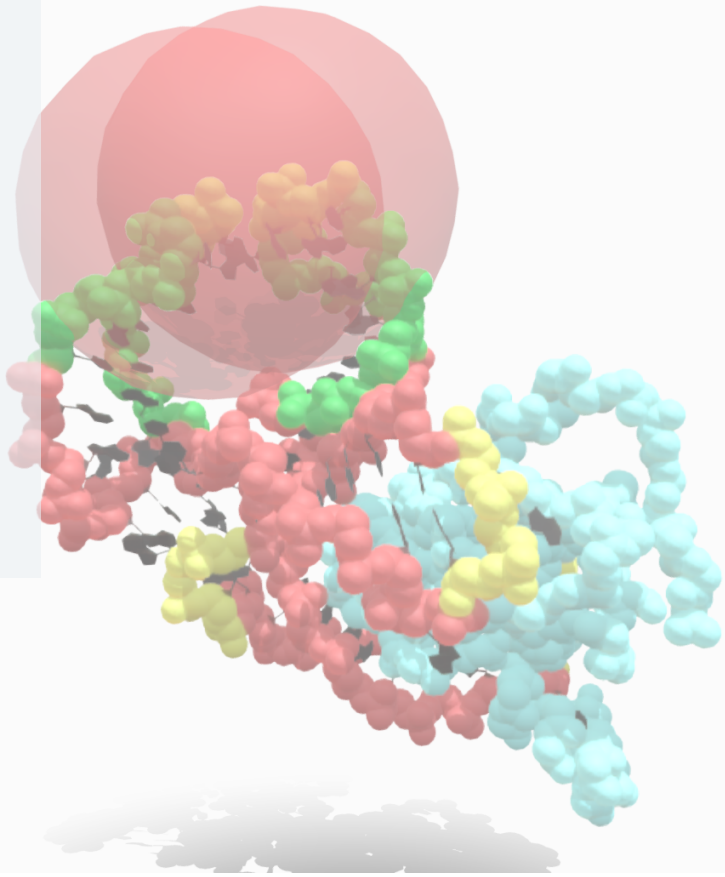


Figura V.15: Ejemplos típicos de la función $f_i(\Delta U)$, definida en la Ecuación V.4 en diferentes rangos de la coordenada de reacción. Los valores de equilibrio para los potenciales sesgados en ventanas adyacentes están indicados por líneas punteadas. Imagen adaptada de la Referencia 3.

Espacio para anotaciones

VI.1 Construcción de modelos 142
VI.2 Simulaciones de *Umbrella Sampling* 142
VI.2.1 Perfiles de Energía Libre 145
VI.3 Mecanismo de unión entre RsmE y sus ARN objetivo 147
VI.4 Racionalización de resultados experimentales previos 148
VI.5 Conclusiones del capítulo 149
VI.6 Anexo 150



VI. Perspectiva: Mecanismo de unión de RsmE a *hcnA*

Esta etapa de investigación se llevó a cabo entre los años 2021 y 2023. Los resultados presentados en este capítulo se encuentran aún en etapa de desarrollo, y trazan las perspectivas de continuidad para esta línea de trabajo.

Resumen de este capítulo En este capítulo se presentan resultados preliminares realizados para dilucidar el mecanismo de unión entre RsmE y uno de sus ARNm blanco. El ARNm estudiado es un fragmento del gen *hcnA*, el cual codifica para una de las subunidades de la enzima responsable de sintetizar HCN en *Pseudomonas protegens*. Dicho metabolito protege a las plantas cercanas frente a patógenos del suelo. El fragmento considerado es un *stem-loop* de tan sólo 20 nucleótidos, con un solo motivo de unión GGA. Por lo tanto, cada molécula de RsmE se une a dos moléculas de ARNm. Los resultados obtenidos con este sistema se comparan con los de otros dos modelos, que consisten en RsmE unida a dos fragmentos de SL2 o dos de SL3. Estudios previos sugieren que SL2 y *hcnA* presentan un carácter anti cooperativo en su unión a RsmE, aunque este efecto no se observa en SL3. Las simulaciones de *Umbrella Sampling*

presentadas en este capítulo reproducen cabalmente los resultados experimentales disponibles para los tres sistemas y, por lo tanto, pueden ser luego aplicados a su racionalización. Al momento de presentarse este escrito nos encontramos trabajando en esta racionalización a fin de establecer criterios para continuar con la línea de trabajo originada con esta tesis.

Los primeros dos capítulos de esta segunda parte de la tesis hicieron referencia al mecanismo molecular a través del cual RsmZ es capaz de unir varias moléculas de RsmE. Además, se estudió el evento de unión entre RsmE y el GGA situado en la porción SC de RsmZ, y se observó que la interacción entre ambas moléculas induce cambios estructurales sobre RsmZ que ajustan su conformación para permitir la unión del siguiente dímero. Lo que resta, por lo tanto, es incluir en el análisis a los ARNm objetivo de RsmE, para alcanzar una descripción global del mecanismo regulatorio abordado en esta tesis. Así, este capítulo busca completar la descripción de los tres principales actores de este sistema de regulación post-transcripcional: La proteína que regula la expresión de un gen a nivel traduccional, el ARNp que impide que ésta ejerza su función, y el ARNm cuya traducción depende de ese tándem. Las siguientes secciones estarán enfocadas en el ARNm que depende de dicho mecanismo.

Como modelo, se tomará la estructura de RsmE unida al constructo "$hcnA_{20-mer}$" mencionado en la Sección I.3, el cual contiene el MU más cercano al codón de inicio de la traducción para la UTR 5' de *hcnA* [26]. Este ARNm codifica para la subunidad A de la enzima HCN Sintetasa, responsable de la síntesis de ácido cianhídrico (HCN) en *Pseudomonas protegens* [307]. Este metabolito secundario provee un mecanismo de protección natural a las plantas cuyas raíces están habitadas por dichas bacterias [64]. Cabe notar que como cada unidad de $hcnA_{20-mer}$ tiene un solo MU, cada unidad de RsmE une dos fragmentos del ARNm. Se ha descrito que la unión de la segunda unidad de $hcnA_{20-mer}$ está desfavorecida respecto a la primera (Ver Sección I.3), pero las razones de este fenómeno son hasta el momento desconocidas.

Con el fin de poder establecer comparaciones significativas, también se incluyó en el análisis a la unión de segmentos de SL2 y SL3 de RsmZ a RsmE. El ingreso de dichos fragmentos a RsmE fue estudiado, junto con el resto de los MU de RsmZ,

en la Referencia 72. Ambos presentan estructura de *stem-loop* con su secuencia A(X)GGAX en el extremo de la horquilla, tal y como ocurre con $hcnA_{20-mer}$. En la referencia 72 se reportó que durante la unión de SL2 a RsmE se observa un aumento significativo del valor de la constante de disociación (Kd) asociada a la segunda unión en relación a la primera [72]. Este comportamiento es análogo al observado con $hcnA_{20-mer}$. Sin embargo, un análisis equivalente realizado sobre el constructo SL3 produjo las mismas Kd para las dos uniones. En conclusión, existen casos en los que la unión de una molécula de ARN a RsmE modifica las condiciones de interacción del SU desocupado de la proteína, y existen casos en los que no. La motivación de este capítulo es abordar dicho aspecto del sistema.

Hipótesis trabajadas en este capítulo

La principal hipótesis evaluada en este capítulo es que, tanto como ocurre con RsmZ, **existe un patrón estructural en RsmE que permite explicar el carácter anti-cooperativo de su unión a *hcnA*$_{20-mer}$** y al constructo SL2. De esta manera, la unión de la primera unidad del ARN induce un cambio conformacional en RsmE que explica el desfavorecimiento en la unión de la segunda molécula. Se asume, por otra parte, que estos cambios son observables por técnicas de DM y *Umbrella Sampling*.

Objetivos de este capítulo

Generales

- Aportar al conocimiento general sobre las bases moleculares del mecanismo de regulación ejercido por las proteínas Csr/Rsm sobre sus ARNm objetivo.
- Racionalizar la información disponible referida al carácter cooperativo o anti-cooperativo de los ARN con capacidad de unión a RsmE.

Específicos

- Comprender las razones de las uniones anti-cooperativas de fragmentos de ARN a RsmE.
- Estimar el costo de energía asociado al primer y segundo evento de unión en los diferentes modelos.
- (Como perspectiva) Identificar a los residuos que ejercen un rol crítico para el proceso de unión.
- (Como perspectiva) Identificar las deformaciones estructurales sufridas por el segundo sitio de RsmE como consecuencia de la primera unión a $hcnA_{20-mer}$.

En las siguientes secciones se detallan los pasos seguidos en esta etapa de la investigación. Las siglas utilizadas en este capítulo son iguales a las de los dos capítulos anteriores, y se encuentran resumidas en la Tabla VII.1.

VI.1 Construcción de modelos

Los modelos estudiados en este capítulo consisten en RsmE unida a fragmentos de ARN en forma de SL, los cuáles contienen secuencias de unión A(X)GGAX en el extremo de la horquilla. Estas secuencias provienen tanto del ARNm *hcnA* como del ARNp RsmZ. El modelo correspondiente a RsmE unido a dos unidades de $hcnA_{20-mer}$ (en adelante RsmE-$(hcnA)_2$) fue construido a partir de la estructura PDB 2JPP [26]. Los modelos que presentan a RsmE unida a los constructos de SL2 o SL3 de RsmZ se generaron con las estructuras reportadas con códigos PDB 2MFE y 2MFF, respectivamente [72]. Dado que cada uno de estos modelos contiene una unidad de RsmE y dos fragmentos idénticos de ARN, serán nombrados como RsmE-$(SLN)_2$, donde N puede ser 2 o 3 según si RsmE está unida a SL2 o SL3, respectivamente.

Partiendo de RsmE-$(hcnA)_2$ y RsmE-$(SLN)_2$, se construyeron modelos que poseen sólo una unidad de su respectivo ARN unido a RsmE. Para esto, se eliminó una de las cadenas de ARN de cada sistema, dando lugar a los modelos RsmE-$(hcnA)_1$ y RsmE-$(SLN)_1$. Esto representa la situación en la que la proteína ya capturó su primera unidad de ARN, mientras su otro SU aún no fue ocupado.

El conjunto de sistemas aquí estudiados representa los pasos consecutivos del mecanismo a través del cual RsmE es capaz de capturar tanto a los fragmentos de RsmZ (SL2 o SL3), como al fragmento $hcnA_{20-mer}$. La Figura VI.1 resume estos mecanismos y cada uno de estos modelos.

VI.2 Simulaciones de *Umbrella Sampling*

Mediante simulaciones de *Umbrella Sampling* se estudió el proceso de salida del constructo $hcnA_{20-mer}$ y de los fragmentos SL2 y SL3 de RsmZ, de los complejos mostrados en la Figura VI.1. La coordenada de reacción se definió de manera

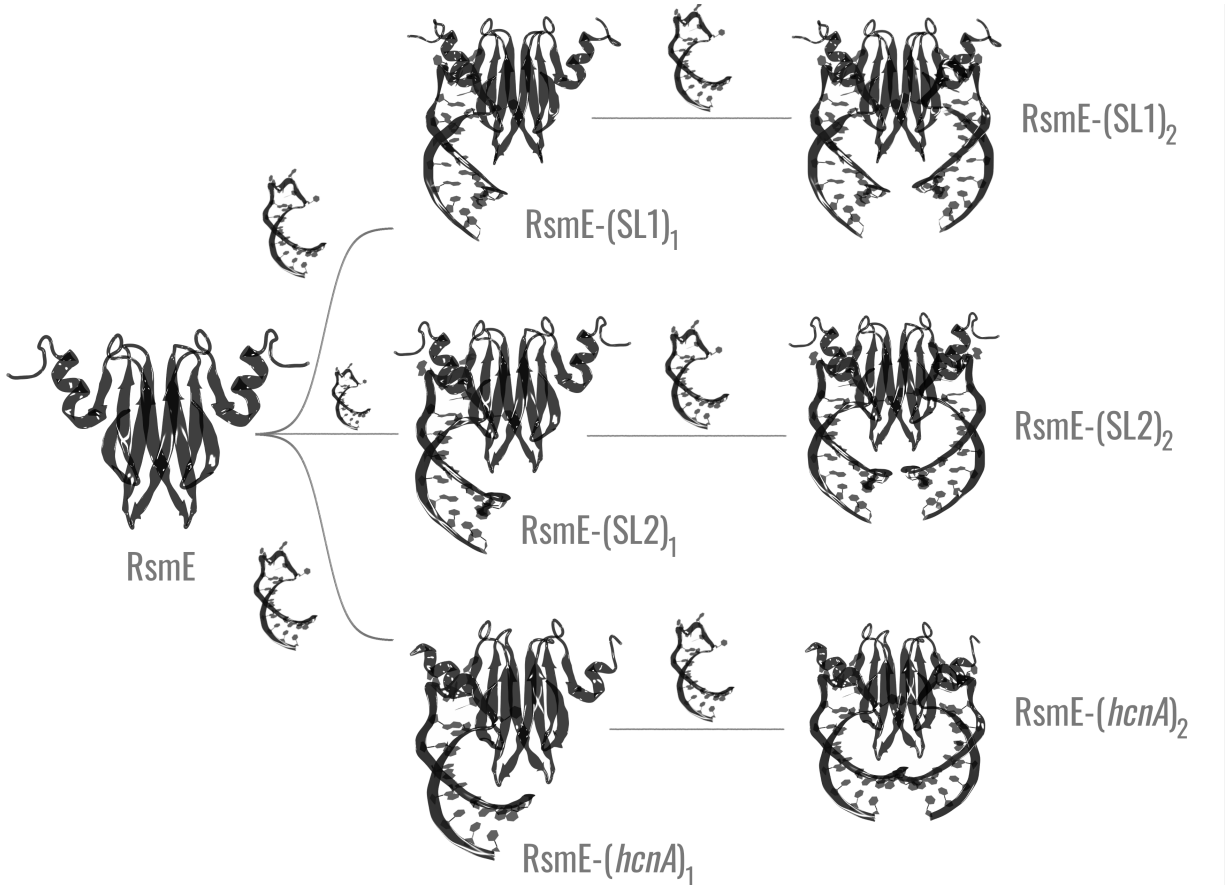


Figura VI.1: Representación esquemática de los modelos estudiados en este capítulo. La Figura representa además, el orden en que las moléculas de ARN se unen a RsmE. De esta manera, los modelos representan los distintos estadios de ocupación de RsmE por parte de SL2, SL3 y $hcnA_{20-mer}$.

análoga a lo descrito en el Capítulo V para estudiar el proceso de unión entre RsmE y la región SC de RsmZ. Así, χ quedó determinada por la distancia entre dos centros de masa: Aquel formado por los átomos de Fósforo de las secuencias específicas de unión de los ARN blanco, y el correspondiente a los C_α de las regiones $\beta 1_A$, $\beta 4_B$ y $\beta 5_B$ de RsmE. En la Figura VI.2 se muestra una representación esquemática de las regiones que definen χ para el modelo RsmE-$(hcnA)_1$. La definición equivalente aplica también a los modelos que presentan a RsmE unida a los fragmentos SL2 y SL3 de RsmZ. Los parámetros utilizados para la simulación fueron los mismos

que los empleados para el Capítulo V. La Tabla VI.1 resume los valores iniciales y finales para la coordenada de reacción χ aplicada sobre cada uno de los modelos que se estudiaron en este capítulo.

Figura VI.2: Representación esquemática para la definición de χ. Los Cα de RsmE se presentan en azul, rojo y naranja dependiendo de si están ubicados en β_1, β_4 o β_5, respectivamente. A su lado, los P de $hcnA_{20-mer}$ están resaltados en verde. En ambos casos, los átomos se representan como esferas, mientras que los esqueletos se muestran como láminas conservando el mismo color.

Modelo	χ inicial (Å)	χ final (Å)
RsmE-(SL2)$_2$	12.07	26.27
RsmE-(SL3)$_2$	13.07	25.17
RsmE-(*hcnA*)$_2$	11.60	35.6
RsmE-(SL2)$_1$	11.26	31.96
RsmE-(SL3)$_1$	12.02	26.72
RsmE-(*hcnA*)$_1$	11.60	35.6

Tabla VI.1: Valores iniciales y finales para las coordenadas de reacción χ utilizadas en cada uno de los modelos estudiados en este capítulo.

VI.2.1 Perfiles de Energía Libre

Tres técnicas alternativas se utilizaron para calcular el PEL de cada evento de separación partiendo de la distribución sesgada de la coordenada de reacción obtenida durante el muestreo. Estas son: WHAM [284], DHAM [285], y BAR [286] (Ver Sección III.4). En todos los casos, se emplearon códigos propios basados en Fortran y los mismos parámetros que los descritos en la Sección V.2. La Figura VI.3 muestra el PEL para la separación entre RsmE y los ARN blanco estudiados en este Capítulo. El panel de la izquierda corresponde a la separación entre RsmE y $hcnA_{20-mer}$, mientras el central y el de la derecha hacen lo propio con SL2 y SL3, respectivamente. Se emplearon, además, varios métodos para corroborar la consistencia de los resultados. Los mismos se encuentran resumidos en el Anexo del capítulo.

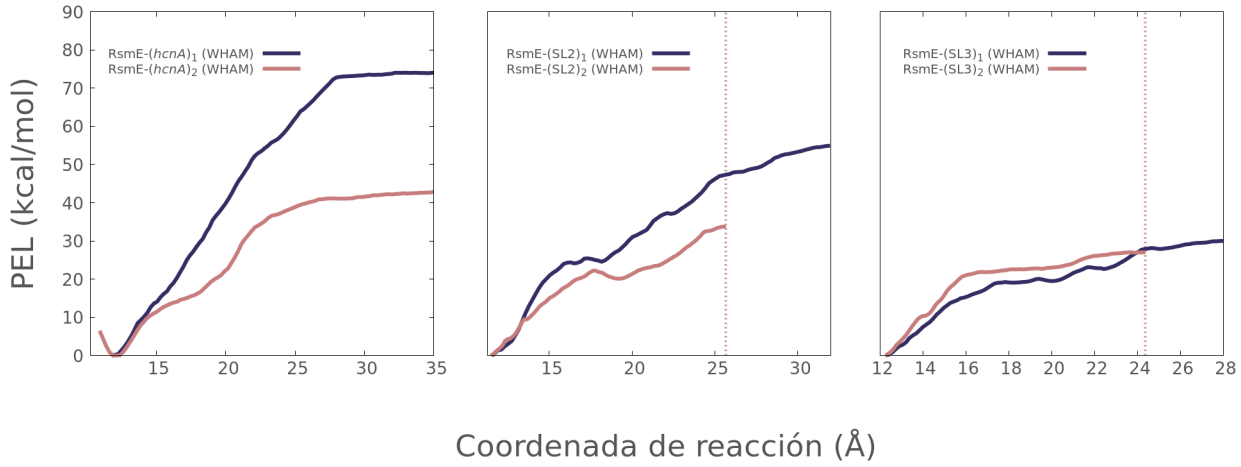


Figura VI.3: PEL para el proceso de separación entre RsmE y tres ARN blancos. Cada curva representa el PEL para la separación entre RsmE y la primera (azul) y segunda (naranja) unidad de ARN. Los resultados de $hcnA_{20-mer}$, SL2 y SL3 se muestran en el panel izquierdo, central y derecho, respectivamente. En estos últimos dos paneles, la línea vertical señala el momento en el que finalizó el proceso de separación para la segunda unidad.

Lo primero que se destaca en la Figura VI.3 es que los eventos de separación de los ARN presentan dos energías de unión disímiles en los complejos formados

por la proteína con $hcnA_{20-mer}$ y SL2. Específicamente, la energía de unión para RsmE-$(hcnA)_2$/RsmE-$(hcnA)_1$ es de ≈-74.0/≈-42.0 kcal/mol, mientras que para el par RsmE-$(SL2)_2$/RsmE-$(SL2)_1$ es de ≈-54.0 kcal/mol /≈-34.0 kcal/mol, respectivamente. Esto indica que el ingreso de la primera unidad tanto de $hcnA_{20-mer}$ como de SL2 es más favorable que el ingreso de la segunda.

Este resultado es coincidente con los estudios experimentales de las Referencias 72 y 73, en las que se reportaron valores de Kd para la disociación de los ARN de los complejos mostrados en la Figura VI.1. Estos valores de Kd se presentan en la Tabla VI.2. Asimismo, cabe destacar que dichos estudios a su vez indicaron que la Kd para el primer evento de unión de SL2 a RsmE es mayor que la de $hcnA_{20-mer}$, indicando una mayor afinidad de éste último por la proteína. Este resultado también es coincidente con nuestras observaciones, donde se ve que el costo de retirar la primera molécula de $hcnA_{20-mer}$ de RsmE-$(hcnA)_1$ es mayor que el de retirar la primera unidad de SL2 de RsmE-$(SL2)_1$.

Por último, para la unión de los fragmentos de SL3, los resultados sugieren que las energías de unión para el primer y segundo evento son similares (≈-30.0 kcal/mol). Esto indica que el carácter anti cooperativo descrito para la interacción de RsmE con $hcnA_{20-mer}$ y SL2, no se observa en este caso. Estos resultados también concuerdan con las mediciones experimentales que muestran idénticas Kd para la unión de la primera y la segunda molécula de SL3 [72].

Motivo	Kd 1ra unión (μM)	Kd 2da unión (μM)	Relación
SL1	1,9	1,9	1,0
SL2	**0,016**	**0,185**	**0,086**
SC	3,5	3,5	1,0
SL3	1,5	1,5	1,0
SL4	2,9	2,9	1,0
hcnA$_{20-mer}$	**0,002**	**0,13**	**0,016**

Tabla VI.2: Valores de Kd para la unión entre RsmE y $hcnA_{20-mer}$/SL2/SL3. Se muestra además la relación entre dichas constantes. Información recopilada de las Referencias 72 y 73.

VI.3 Mecanismo de unión entre RsmE y sus ARN objetivo

La separación entre la primera unidad de $hcnA_{20-mer}$ y RsmE puede verse en la animación disponible en https://www.youtube.com/watch?v=Nfti6mAUo1c. Lo primero que puede notarse al analizar la animación en el sentido inverso (y describir, por lo tanto, el proceso de unión) es que el *loop* que contiene a la secuencia ACGGAU es la primera porción del ARN en interactuar con RsmE. Este hexanucleótido adopta una forma curva con la guanosina central (ACG**G**AU) funcionando como vértice del motivo. Este comportamiento es análogo al observado durante la unión del motivo SC_{GGA} de RsmZ a RsmE descripta en el Capítulo V. El mismo comportamiento se encuentra también al mirar, en sentido inverso, la unión de una segunda molécula de $hcnA_{20-mer}$ al complejo RsmE-$(hcnA)_1$, disponible en https://www.youtube.com/watch?v=kbE9bEFJnC0.

Las animaciones correspondientes a los eventos de separación de los modelos RsmE-$(SL2)_1$ y RsmE-$(SL2)_2$ se encuentran disponibles en los enlaces https://www.youtube.com/watch?v=dtrzWI9huao y https://www.youtube.com/watch?v=6FLw-zGOEZI, respectivamente. Las consideraciones mencionadas en el párrafo precedente aplican también a este sistema. Las regiones de doble cadena de SL2 son las primeras en separarse de la proteína en ambos modelos, mientras los motivos AAGGAU son los últimos. Finalmente, los enlaces https://www.youtube.com/watch?v=DK1rOdU9jyY y https://www.youtube.com/watch?v=DbfBKvbnYoU presentan las animaciones correspondientes a los modelos RsmE-$(SL3)_1$ y RsmE-$(SL3)_2$ durante el proceso de separación del complejo. Puede comprobarse que, cualitativamente, las características de estos eventos coinciden con la de los mencionados anteriormente.

Como conclusión preliminar, las simulaciones de *Umbrella Sampling* predicen el carácter anti cooperativo para la unión de RsmE tanto a $hcnA_{20-mer}$ como al segmento SL2 de RsmZ. Asimismo, indican que este efecto no existe en el caso de los complejos formados con SL3. Al momento de presentar esta tesis nos encontramos analizando los resultados de las simulaciones en busca de una explicación plausible para este comportamiento. Los párrafos siguientes buscan racionalizar los datos experimentales con los resultados preliminares recién presentados. A partir de esta

racionalización se pretende establecer nuevas conjeturas sobre las razones para el carácter anti cooperativo de unión entre RsmE y los constructos $hcnA_{20-mer}$ y SL2, sentando las bases sobre las cuáles diseñar futuros análisis para comprobar estas hipótesis.

VI.4 Racionalización de resultados experimentales previos

La Referencia 72 discute el carácter anti cooperativo de la interacción entre $hcnA_{20-mer}$ y RsmE a partir de valores de Kd para las uniones correspondientes a la primera y segunda molécula del ARN a la proteína. El mismo análisis se da sobre los diferentes MU de RsmZ de manera independiente (SL1, SL2, SC, SL3, y SL4). Sólo el constructo $hcnA_{20-mer}$ y el SL2 de RsmZ presentan dos valores diferentes de Kd para cada uno de los eventos de unión. El resto de los MU de RsmZ presentan sólo una constante. Estas observaciones están resumidas en la Tabla VI.2.

Además, $hcnA_{20-mer}$ y el fragmento de SL2 presentan los menores valores de Kd de todo el conjunto. Esto sugiere que podría existir una relación entre la alta afinidad de $hcnA_{20-mer}$ y SL2 por RsmE, y el carácter anti cooperativo de su mecanismo de unión. Respecto a esto, cabe mencionar que la Referencia 72 destaca que modificar la secuencia A(X)GGAX de los motivos de unión implica cambios significativos en la afinidad por RsmE. En particular, el trabajo postula que la omisión del nucleótido "(X)" anterior al motivo GGA implica variaciones de entre 10 a 1000 veces en la afinidad del ARN por RsmE. Este es el caso de SL3 y SL4, los cuáles a diferencia de SL2 y el constructo $hcnA_{20-mer}$, no poseen dicho residuo. SL1 y SC, en cambio, tienen una extensión subóptima de la región simple cadena sobre la cual se ubica el motivo GGA que explicaría su menor afinidad por RsmE [72].

Tomando en consideración lo discutido por la Referencia 72, y lo sugerido por nuestros resultados, la hipótesis que se plantea a modo de perspectiva es que la presencia del nucleótido "(X)" en SL2 y $hcnA_{20-mer}$ contribuye a las deformaciones sufridas por RsmE como consecuencia del primer evento de unión. De esta manera, las interacciones formadas por dicho nucleótido tienen un efecto sobre el mecanismo de unión que no se encuentra presente para el caso de SL3. Futuros ensayos enfocados en el rol de dicho residuo deberán implementarse para poner a prueba

dicha hipótesis.

A modo de hipótesis a nivel biológico, se propone que este mecanismo de anti cooperatividad en contextos de alta afinidad de RsmE por sus ARN objetivo podría funcionar como una forma de autorregular su propia acción sobre dichos ARN. De esta manera, se ve compensada su alta afinidad con una menor predisposición a capturar más de una unidad de dicha molécula.

VI.5 Conclusiones del capítulo

En este capítulo se estudió el mecanismo de reconocimiento y unión entre RsmE y un segmento del ARNm *hcnA* correspondiente a parte de su UTR 5' ($hcnA_{20-mer}$). A su vez, se analizó el mismo fenómeno entre dicha proteína y los segmentos SL2 y SL3 de RsmZ. Para ello, se implementaron cálculos de *Umbrella Sampling* con el fin de explorar el proceso de unión y estimar su costo de Energía Libre asociado. Hasta la fecha, estas fueron las primeras simulaciones empleadas para estudiar estos procesos.

Los resultados concuerdan con el mecanismo anti cooperativo de unión entre RsmE y $hcnA_{20-mer}$ previamente reportado en estudios experimentales. Así, la unión del primer segmento del ARNm está más favorecida que la del segundo al observar sus respectivos PEL. Este mismo comportamiento fue observado también para la unión entre RsmE y SL2. No obstante, la unión entre RsmE y SL3 no muestra diferencias significativas entre el primer y segundo evento, indicando la ausencia de un carácter anti cooperativo en este modelo. Estas observaciones también coinciden con lo reportado de manera experimental.

Estos resultados demuestran que los cálculos de *Umbrella Sampling* son adecuados para captar las razones moleculares de la anti cooperatividad de la unión de $hcnA_{20-mer}$ y SL2 a RsmE, y distinguir su comportamiento de la unión a SL3. Sin embargo, resta aún identificar cuáles son las causas de este comportamiento. Cabe recordar en este punto que la principal diferencia entre SL3 y los otros dos constructos es la presencia de un nucleótido adicional en el segmento de simple cadena en el extremo de la horquilla. A modo de perspectiva, se propone explorar la influencia que dicho nucleótido tiene sobre la estructura de la proteína al unirse

la primera unidad de los ARN que lo contienen.

En relación a la hipótesis planteada al principio del Capítulo, se refuerza la conjetura de que existe un patrón que permite explicar el carácter anti cooperativo para la unión de RsmE a $hcnA_{20-mer}$ y al constructo SL2. Este patrón evidentemente depende de la fuerza con la que la proteína une a la molécula de ARN, que a su vez viene dada por la secuencia de esta última. El entendimiento de este mecanismo particular de funcionamiento podría tener implicaciones en el diseño de alternativas de regulación bacteriana a nivel post-transcripcional. No obstante, resulta necesario establecer las bases moleculares de dicho comportamiento, planteándose así una perspectiva de continuidad para la línea de investigación comenzada con este trabajo.

VI.6 Anexo

El primer ensayo aplicado para evaluar la consistencia de los cálculos de *Umbrella Sampling* consistió en comparar los valores obtenidos a partir de los métodos de WHAM, DHAM y BAR. Las curvas correspondientes se muestran en la Figura VI.4.

Adicionalmente, se aplicó el test de Kullback-Leibler (Ver Ecuación V.2), que mide la similitud entre dos distribuciones. En este caso, se considera cada ventana del cálculo y se toma la distribución sesgada muestreada directamente. Al mismo tiempo, se construye la distribución obtenida luego de aplicar nuevamente el sesgo a la distribución de datos desesgados por WHAM. Estas dos distribuciones deben ser similares en caso que el muestreo se encuentre convergido. A menor valor de S, mayor semejanza entre las distribuciones (Ver Ecuación V.2). La Figura VI.5 muestra los resultados de este análisis para los modelos estudiados en el presente capítulo.

La Figura VI.6 muestra los resultados del ensayo definido por la Ecuación V.4 para los modelos RsmE-$(hcnA)_1$ y RsmE-$(hcnA)_2$, donde se obtienen las diferencia de Energía libre calculadas a partir de los datos solapados entre ventanas adyacentes del muestreo. En todos los casos, se seleccionaron pares de ventanas representativas del comienzo (panel izquierdo), la mitad (panel central), y el final (panel derecho)

del rango de la coordenada de reacción. Las Figuras VI.7 y VI.8 muestran el equivalente a la Figura VI.6, para los complejos formados por RsmE con SL2 y SL3, respectivamente.

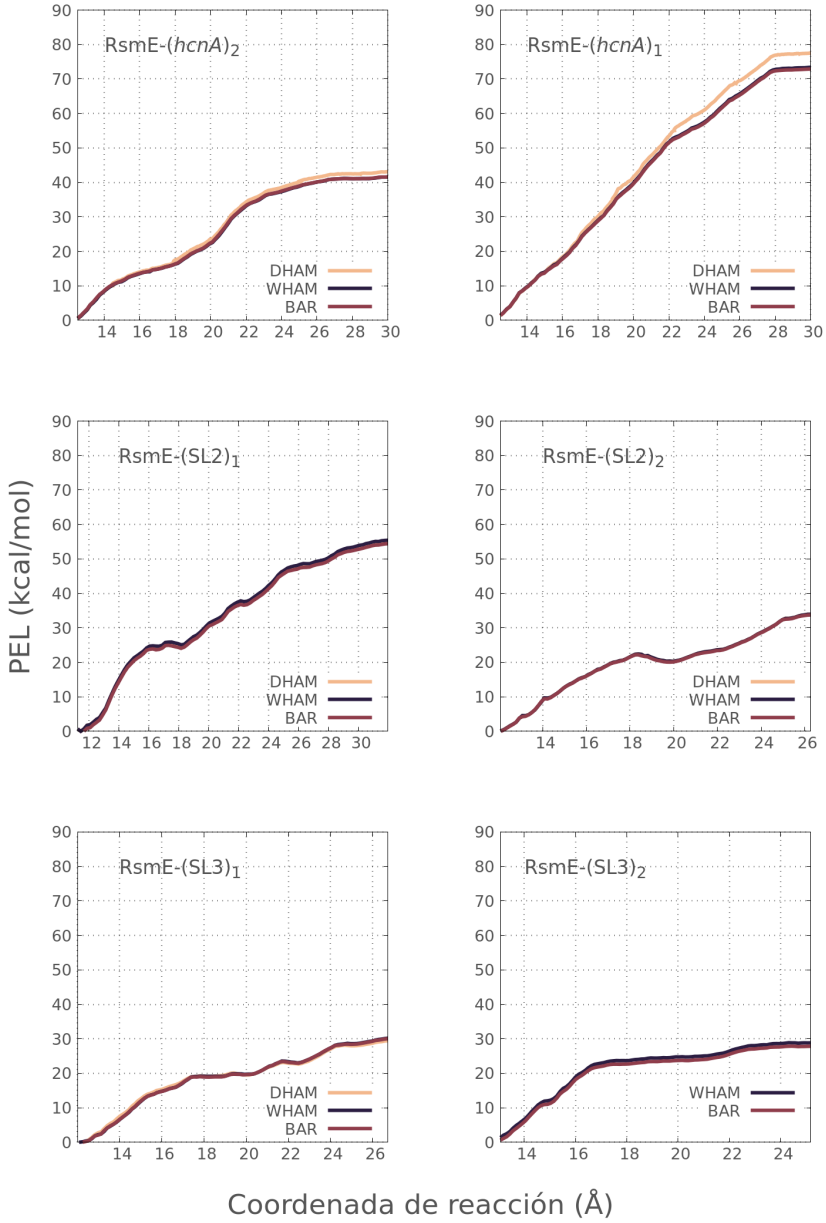


Figura VI.4: Comparación de los PEL calculados con WHAM (azul), DHAM (naranja) y BAR (rojo) para los seis modelos estudiados en este capítulo.

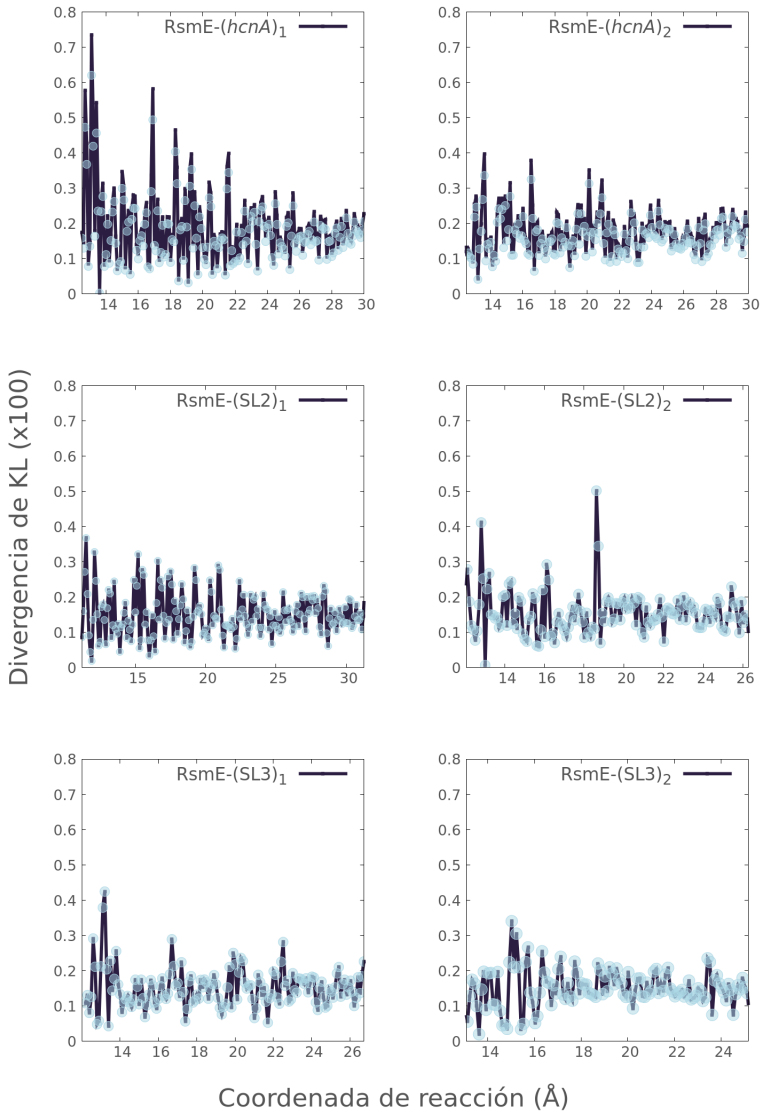


Figura VI.5: Divergencia simétrica de Kullback-Leibler (S_i), definida en la Ecuación V.2, en función de la coordenada de reacción χ para los seis modelos estudiados en este capítulo. Todos los valores de (S_i) están por debajo de 0,0073, lo que indica una coherencia razonable entre las distribuciones sesgadas y desesgadas.

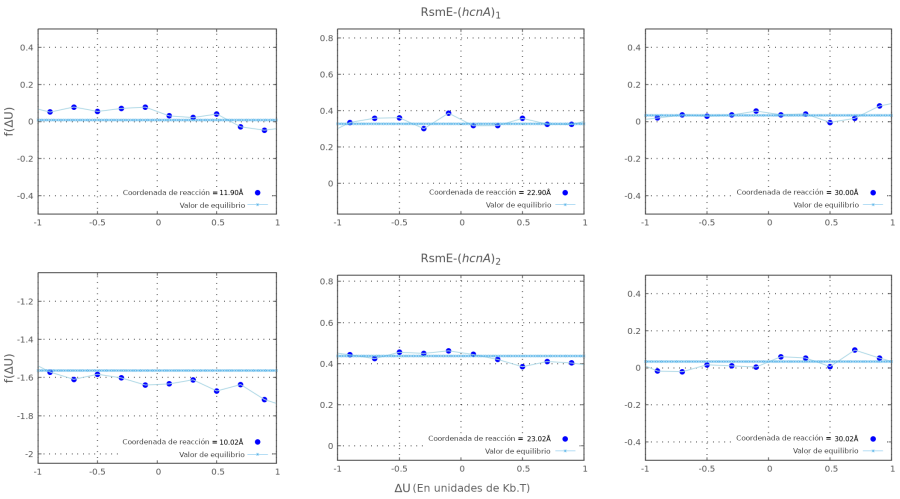


Figura VI.6: Ejemplos típicos de la función f1(ΔU) - f0(ΔU), definida en la Ecuación V.4, en tres rangos diferentes de coordenadas de reacción para RsmE-(*hcnA*)$_1$ (panel superior) y RsmE-(*hcnA*)$_2$ (panel inferior).

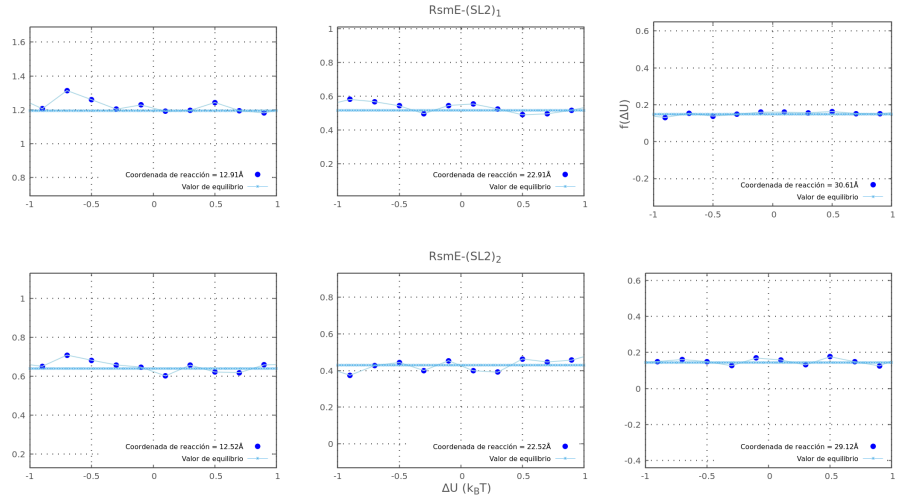


Figura VI.7: Ejemplos típicos de la función f1(ΔU) - f0(ΔU), definida en la Ecuación V.4, en tres rangos diferentes de coordenadas de reacción para RsmE-(SL2)$_1$ (panel superior) y RsmE-(SL2)$_2$ (panel inferior).

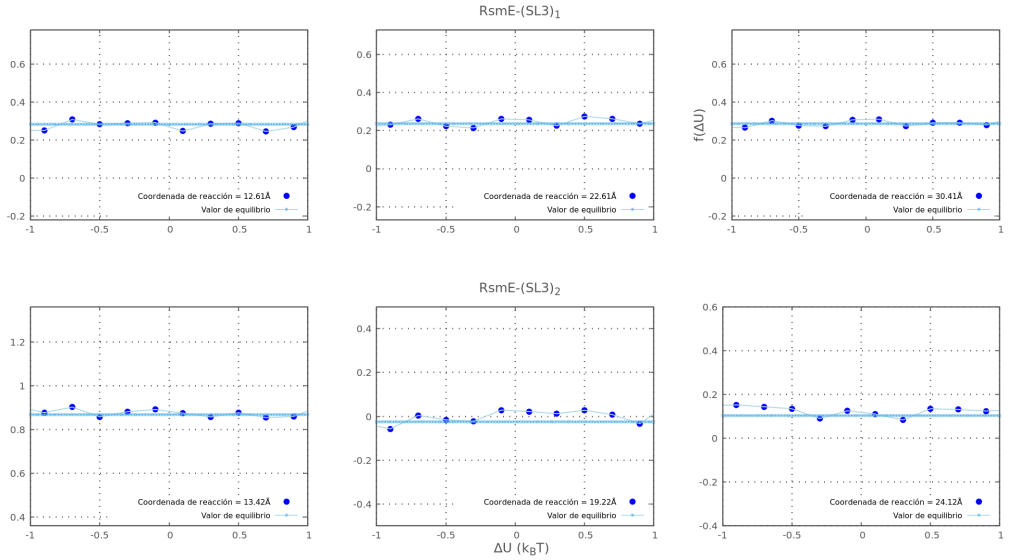
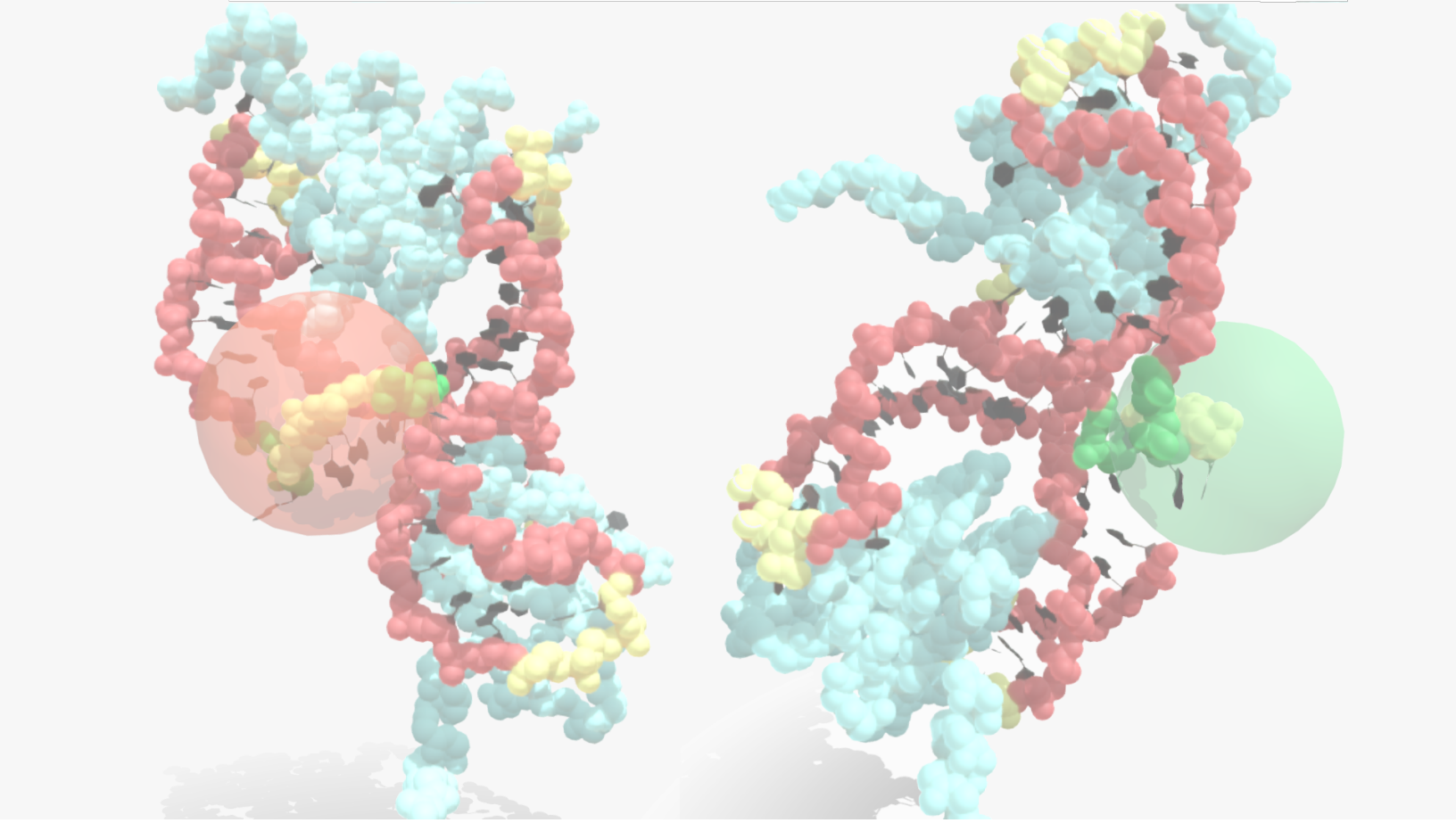


Figura VI.8: Ejemplos típicos de la función f1(ΔU) - f0(ΔU), definida en la Ecuación V.4, en tres rangos diferentes de coordenadas de reacción para RsmE-(SL3)$_1$ (panel superior) y RsmE-(SL3)$_2$ (panel inferior).

Espacio para anotaciones



VII. Discusión y Conclusiones generales

La estructura de esta tesis tuvo hasta este punto un orden que permite a cada capítulo nutrirse de los conceptos introducidos en sus predecesores para hacer uso de ellos al momento de abordar nuevas consideraciones. En este último apartado, sin embargo, la idea será brindar una perspectiva integral del conjunto de los temas tratados y discutir sus implicancias para comprender el mecanismo de regulación post-transcripcional mediado por RsmE.

Esta tesis comenzó con una primera parte cuyo objetivo fue introducir las bases teóricas sobre las que fueron construidas las diferentes etapas de la investigación que luego serían descritas en la segunda parte del escrito. En el Capítulo II se discutieron las características estructurales de los ARN, enfatizando el rol biológico de los *stem-loops* y los empalmes. Los resultados de los Capítulos IV, V y VI demuestran que estos motivos estructurales son clave para comprender la interacción de RsmZ

y *hcnA* con RsmE. El Capítulo V demuestra, además, que el rol asumido por los empalmes en su interacción con RsmE depende de su longitud. De esta manera, se requiere una longitud mínima que le permita adoptar una estructura simil *stem-loop* para lograr unir a RsmE.

Asimismo, las interacciones entre bases desapareadas en los extremos de *stem-loops*, así como motivos del tipo *ribose zipper* presentados en el Capítulo II, fueron identificados como influyentes sobre los estados de disponibilidad/oclusión discutidos en los modelos explorados en el Capítulo IV. Estos motivos ya habían sido identificados por posibilitar la alternancia entre estados activos e inactivos de los ARN [152], de manera que los resultados aquí presentados concuerdan con estas observaciones. Cabe destacar que mientras estas interacciones terciarias fueron observadas en determinados fragmentos de las simulaciones realizadas, los motivos secundarios permanecieron prácticamente inalterados en todo momento. Esto da cuenta de la estabilidad de los modelos para las condiciones de simulación empleadas, lo cual demuestra que los campos de fuerza seleccionados funcionaron adecuadamente [91]. Cabe recordar que, como se discutió en el Capítulo II, la disrupción de este tipo de interacciones requiere superar altas barreras de Energía Libre, y por lo tanto, ocurre en tiempos que exceden a los alcances de la DM clásica [153]. Así, los eventos observados en los tiempos de simulación que aquí se emplearon están dentro de los órdenes esperables por bibliografía.

En relación a esto último, debemos indicar que las conformaciones L y R de RsmZ parecen representar dos mínimos globales separados por altas barreras a lo largo de la SEL de RsmZ. Incluso en los tiempos de simulación empleados en este trabajo, los cuales son suficientes para alcanzar resultados consistentes respecto a otras observaciones experimentales, no se encontró evidencia de una interconversión entre ambas conformaciones. Esto sugiere que ambas coexisten en cuasi-equilibrio, reforzando la hipótesis planteada por la Referencia 73 respecto a la importancia de que tanto el confórmero L como el R estén presentes en el medio intracelular para adaptar su mecanismo de acción a distintos contextos de unión. Sin embargo, para poner a prueba esta hipótesis, deberían implementarse métodos de muestreo mejorado que permitan determinar el camino seguido por este cambio

conformacional, así como estimar su costo energético.

Como se mencionó en el Capítulo III, las simulaciones computacionales sobre ARN, tanto como de métodos de muestreo mejorado, son aún infrecuentes en la bibliografía sobre este tipo de moléculas. En este sentido, los cómputos presentados en este escrito se suman a los escuetos antecedentes sobre abordajes computacionales abocados a la simulación de ARN pequeños y no codificantes (ARNp), como es el caso de RsmZ. Pero además, implican las primeras simulaciones reportadas para el complejo formado entre dicho ARNp y RsmE.

Ya refiriéndome concretamente al sistema RsmE-RsmZ, los resultados proporcionados en esta tesis exploran aspectos estructurales y dinámicos que previamente no habían sido descritos, y suman observaciones que permiten racionalizar resultados experimentales previos sobre los que se desconocían sus bases moleculares. En el Capítulo IV se presentan estructuras típicas para RsmZ en distintos estados de ocupación, las cuáles no estaban hasta el momento discutidas. No obstante, es importante no perder de vista que dichas estructuras se tratan, ante todo, de un modelo. Por lo tanto, es necesaria la complementación con estudios experimentales que partan de las hipótesis sugeridas por estos modelos y las pongan a prueba.

Los cálculos del Capítulo IV revelaron un patrón estructural y dinámico en RsmZ que explica el carácter secuencial y cooperativo de su unión a diferentes unidades de RsmE. Dicha secuencia fue racionalizada observando el orden en el que se exponen al medio los motivos de unión del primer, segundo y tercer bolsillo descritos en las estructuras reportadas por la Referencia 73. En este sentido, es interesante destacar que en el año 2023 fueron reportadas dos nuevas estructuras para la secuencia completa de RsmZ [78], las cuales coinciden con las presentadas en la Referencia 73 respecto a que los *stem-loops* 2 y 3 (SL2 y SL3, respectivamente) forman el primer bolsillo en ser ocupado por RsmE. Sin embargo, difieren en cómo continúa la secuencia de unión. En el Capítulo IV se discute cómo ésto pone en evidencia la influencia que tiene la longitud del modelo sobre los resultados obtenidos. Pero además, propone que la alternancia observada en nuestras simulaciones respecto al nivel de exposición de SL4 podría explicar la discrepancia entre ambos estudios experimentales.

Otro de los resultados relevantes de esta tesis es que los segmentos de simple cadena que contienen motivos de unión para RsmE pueden adoptar estructuras de símil *stem-loop* para interaccionar con dicha proteína cuando poseen la longitud adecuada. Esto se deduce de los Capítulos IV y V. En el primero de ellos se observa que hay una interacción entre los residuos C37 y A41 del segmento monocatenario que empalma a SL2 y SL3. La misma sólo está presente en las estructuras en las que el MU ubicado en ese segmento aparece como accesible para la unión de RsmE. Esta interacción le permite al segmento adoptar una forma de vértice hacia afuera de la molécula, en la que el residuo G_{40} es el más expuesto. Por su parte, en el Capítulo V se destacó que la adopción de esa conformación ocurre también durante el proceso de su unión a RsmE.

Este fenómeno puede ser relevante porque favorece el reconocimiento entre cada parte del tándem RsmE-RsmZ, donde los residuos participantes de las primeras etapas del evento quedan mejor dispuestos para su reconocimiento. De hecho, los ensayos empleados en esta tesis, que combinan tanto mapas de contactos como Análisis de Componentes Principales (PCA), muestran que los residuos que participan en las primeras etapas del evento de unión coinciden con los que fueron previamente reportados en bibliografía como los más influyentes para el comportamiento normal del sistema [73, 79, 80]. G_{40} es, precisamente, uno de esos residuos.

En este sentido, es interesante destacar que aunque RsmE y RsmZ son capaces de adoptar de manera libre conformaciones similares a las formadas durante el proceso de unión entre estas, las mismas son alcanzadas de forma marginal en ausencia de su contraparte. De esta manera, si bien la mayor flexibilidad de cada una de estas moléculas está precisamente concentrada en las regiones que participan del proceso de unión, la presencia de la molécula con la cual interactúan es necesaria para que terminen de adoptar la conformación óptima necesaria para que el evento se produzca. Si bien en trabajos previos se sugirió que dicha relevancia se debe a la afinidad secuencia-específica entre las partes, en este caso se suma además un rol adicional asignado a esos residuos (particularmente, el nucleótido central de los motivos GGA en RsmZ y *hcnA*, así como del residuo Lys38 en RsmE), relacionado

con su carácter dinámico previo al proceso de unión.

En el Capítulo VI se estudió la unión de RsmE a diferentes fragmentos de sus ARN objetivo y se encontró que los resultados de simulaciones de *Umbrella Sampling* concuerdan con las observaciones experimentales. Así, la unión de dos fragmentos del constructo $hcnA_{20-mer}$ o dos de SL2 presentan un carácter anti cooperativo, en el sentido de que la segunda unión es mucho más débil que la primera. Sin embargo, en el caso de la unión a dos fragmentos de SL3, las dos uniones tienen estabilidades similares. La concordancia entre estas observaciones y los resultados experimentales reportados en las Referencias 72 y 73 es destacable. Sin embargo, resta dilucidar las razones de este comportamiento.

La hipótesis que se postula es que cuando RsmE en estado libre se une a un MU de elevada afinidad, como los que están en los *loops* de $hcnA_{20-mer}$ o SL2, se produce una distorsión en su estructura que altera su capacidad de interaccionar con un segundo MU. Esto no ocurre cuando RsmE se une a motivos de baja afinidad, como los del *loop* de SL3. De esta manera, la alta afinidad de RsmE por un dado ARN podría verse compensada por una menor tendencia a capturar una segunda unidad de dicha molécula. Dicho comportamiento podría sumarse a otras estrategias de autorregulación postuladas para RsmE en el Capítulo I. Al momento de escribir esta tesis, estamos diseñando los experimentos para poner a prueba esta hipótesis. Para esto se propone como perspectiva poner el foco en cómo influye sobre dichas transformaciones la presencia o ausencia del nucleótido "(X)" en la secuencia A(X)GGAX, dada la influencia comprobada que éste tiene sobre la afinidad entre RsmE y sus ARN objetivo [72].

Si bien en su forma aislada SL2 presenta un comportamiento anti cooperativo para su unión a RsmE, de manera que la unión de una segunda unidad de este motivo se ve desfavorecida respecto a la primera [72], está descrito que SL3 aumenta su afinidad por RsmE cuando está covalentemente unido a SL2 en la molécula completa respecto a cuando se lo estudia de forma individual [73]. Teniendo en cuenta que SL2 es el MU de RsmZ que reporta la mayor afinidad por RsmE [72], y que según nuestras observaciones en el Capítulo IV presenta una alta accesibilidad cuando RsmZ se encuentra de forma libre, es esperable que sea uno de los primeros en unir

a la proteína. Teniendo en cuenta que SL3 está inmediatamente unido a este motivo, es de esperar que se vea favorecida su posterior unión a RsmE tras el ingreso de SL2 por quedar dispuesto de manera cercana al sitio de unión restante de la proteína. Esto podría explicar que las Referencias 73 y 78 coinciden en que el par SL2-SL3 es el primero en ser ocupado por RsmE.

Finalmente, se plantea como perspectiva el modelado de porciones mayores de *hcnA*, partiendo de su secuencia conocida. En el Capítulo III se discutió que los algoritmos que apelan a herramientas basadas en IA y Aprendizaje Automatizado para la predicción de estructuras 3D de ARN basadas en su secuencia aún no generan resultados tan satisfactorios como las que hacen predicciones sobre proteínas. Por su parte, el Capítulo II enfatiza la importancia de caracterizar al ARN no a partir de una única estructura, sino de un ensamble de conformaciones. Debido a que se trata éste de un campo en constante crecimiento con algunas metodologías prometedoras actualmente en desarrollo [308], sería interesante ponerlas prueba para poder modelar porciones mayores de *hcnA* a fin de describir su diversidad conformacional y la exposición/oclusión de sus numerosos MU, de manera análoga a como se hizo con RsmZ.

Los principales resultados surgidos de esta tesis, así como la racionalización antes mencionada para los resultados experimentales previos, pueden sistematizarse en un único modelo que describe cómo se produce la unión entre RsmE y cualquiera de los ARN capaces de interactuar con dicha proteína reportados a la fecha. Este modelo se puede resumir de la siguiente manera:

- La capacidad de un motivo dado para unir a RsmE está determinada tanto por su disposición en una región de simple cadena, como de su accesibilidad medida en términos de exposición al solvente [2, 72].
- En aquellos casos donde la región simple cadena no se encuentre en el extremo de un *stem-loop*, sino en un empalme, éste deberá tener una longitud adecuada para poder adoptar una conformación similar al de una horquilla durante su unión a RsmE [3].
- En aquellos ARN cuya afinidad por RsmE sea baja, el mecanismo de unión estará determinado por la identidad del residuo X en la secuencia A(X)GGAX,

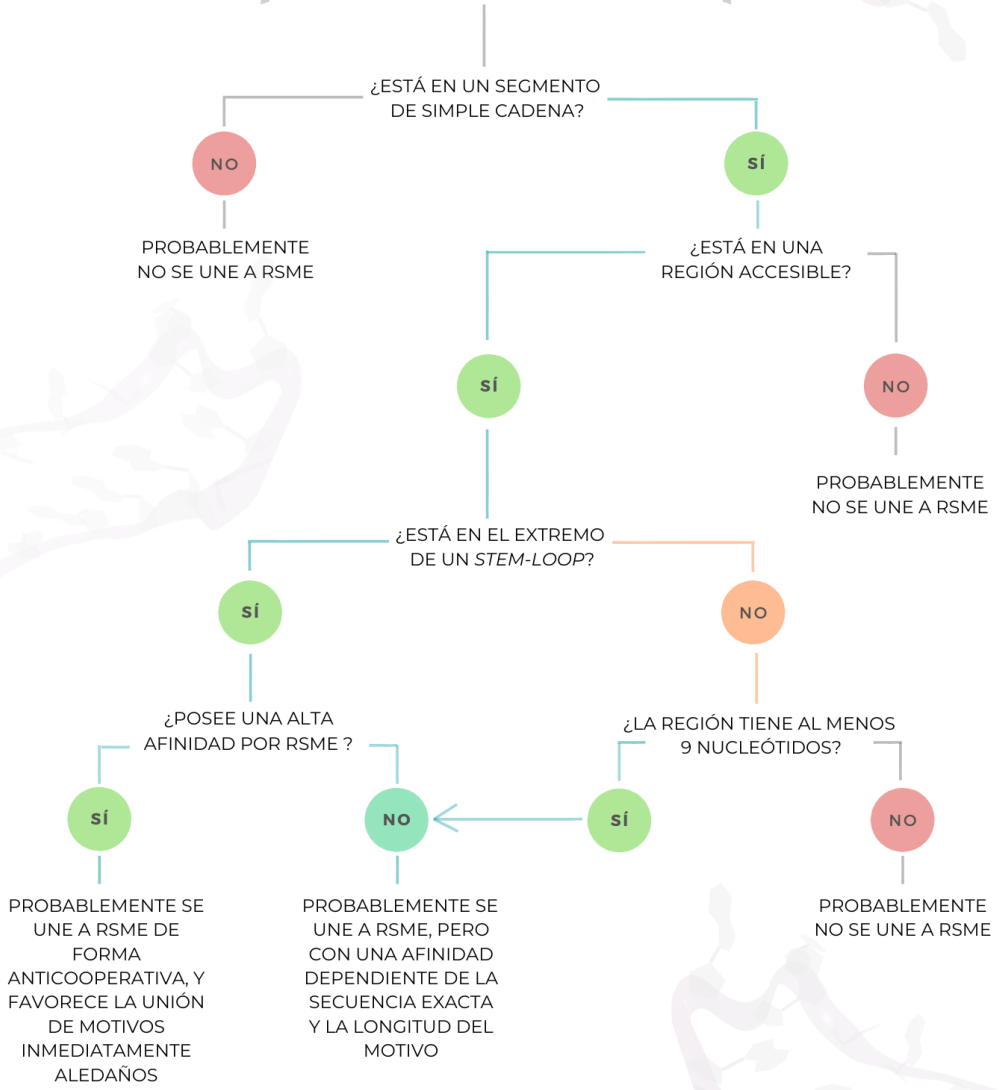
así como por su presencia o ausencia en la posición "(X)" [72].

- Por el contrario, si el motivo en cuestión posee una alta afinidad por RsmE, la unión de fragmentos independientes (no unidos entre sí) a la proteína seguirá un mecanismo anti cooperativo, donde la unión de la segunda unidad estará desfavorecida con respecto a la primera [72, 307].
- En el caso de que el motivo mencionado en la opción anterior esté inmediatamente unido a otro, éste último se verá favorecido a ocupar el segundo sitio de RsmE por cuestiones de proximidad. Esto facilita su accesibilidad por sobre motivos de una región distante de la molécula, o bien motivos independientes que no están unidos directamente al primero en ingresar.

La figura de la página siguiente resume estos postulados con un diagrama de decisión. De esta manera, esta tesis propone un modelo que sistematiza y racionaliza el conocimiento estructural y dinámico sobre la familia de proteínas Csr/Rsm, y cómo se produce su interacción con los ARN capaces de unirse a ellas.

Espero que esta contribución, más allá de los alcances personales que implicó el hecho de haberla llevado a cabo, haga su aporte al estado del arte tanto del estudio del ARN mediante métodos computacionales como del sistema de regulación post-transcripcional mediado por el tándem RsmE-RsmZ. En particular, espero que el conocimiento así generado pueda ser retomado por estudios posteriores para el desarrollo de tecnologías abocadas a tener una agricultura sustentable y amigable con el medio ambiente, haciendo uso de las características fenotípicas de las bacterias que contienen al sistema estudiado, y del impacto directo que éste tiene sobre la síntesis de metabolitos con potencial biorregulador.

CÓMO SERÁ LA UNIÓN DE UN MOTIVO A(X)GGAX A RSME



VII.1 Agradecimientos

Vengo de una familia donde todo el mundo venía de otra parte. Por parte de madre, inmigrantes que llegaron a nuestro país escapando de la posguerra italiana, y la Guerra Civil Española. Por parte de padre, gente del interior que descendía a su vez de personas que no pudieron quedarse en su país de origen. Toda esa gente se encontró fortuitamente en el Conurbano Bonaerense, y sus hijes se encontraron a su vez en la educación pública. A esa familia, a ese Conurbano, y a esa educación pública es a quienes les dirijo mi primer agradecimiento, porque son el punto de inicio de mi propia historia.

Soy hijo de esa familia, pero también de *lo público*; Ese significante en disputa tan propio de la época en la que fueron escritas estas palabras. Por lo tanto no quiero ahorrar en agradecimientos a los Jardines de Infantes 901 y 915; a la Escuela Primaria N°30 "Manuel Belgrano"; A la Escuela Técnica N°2, "Paula Albarracín de Sarmiento" ("El Chaparral", para les amigues); y muy especialmente, al CONICET, y a la Universidad Nacional de Quilmes, que merece un párrafo aparte. Tal vez dos.

Me tocó vivir, desde que inicié mis estudios en Biotecnología en el 2012, la pérdida sucesiva de quienes integraban esa familia que me dió origen. Ante la disrupción de ese hogar lleno de acentos raros, la UNQ pasó a ser el nuevo hogar. Es una frase hecha decir que lugares amados como este son "la segunda casa", pero para mi la UNQ es la primera. No hay lugar en el que me sienta más en casa que en la UNQ. Una casa que por supuesto no me es propia, y creo que en definitiva es esa la mejor parte: que es de la comunidad.

La UNQ y Quilmes se parecen mucho en un aspecto, y es que si bien no son pocas las personas que forman parte de su comunidad, se conserva una estructura de barrio que nos lleva a conocernos más o menos entre todes. Dentro del submundo universitario argentino, la UNQ es una especie de barrio chico, con sus costumbres y berretines. La gente de Quilmes es muy de Quilmes, localista, fundamentalista, fanática en el mejor sentido del fanatismo, que es el del orgullo. Algo parecido pasa con nuestra Universidad. Si no existiera, habría que inventarla.

Quilmes es un lugar particular porque produce la cerveza más popular del país, el alfajor con más mística del que se tenga memoria, y también produce Ciencia.

No es poca cosa, porque no es un lugar del que se espere algo así. Y cuando digo "que se espere" me refiero a las expectativas que tienen de un barrio como el nuestro quienes fijan el cartograma de lo que debería hacer (o dejar de hacer) cada rincón de mundo: Quilmes queda muy lejos de los lugares en los que se supone que uno puede ser feliz, y sin embargo lo intentamos igual. Y a veces, hasta lo logramos. Hacer Ciencia desde Quilmes es una decisión política, casi rebelde. Es a Quilmes, a su gente, y a su Ciencia a quienes les quiero dedicar un agradecimiento en particular también.

Lo otro que me pasa con Quilmes es que también es área de concentración para la mayoría de las cosas que amo. No es sólo el olor a malta a la mañana en Vicente López y 12 de Octubre, ni la hora dorada en la Ribera, o el otoño en la Villa Argentina. No es sólo la EMBA, la peatonal Rivadavia, o sus dos clubes de fútbol. Ni siquiera es sólo la UNQ. Es que la mayor parte de mi familia elegida, esa que me fui armando a medida que se fue yendo la mía de origen, está en Quilmes. Por supuesto que hay notables excepciones, que traigo a colación para no recaer en el dogmatismo (y sobre todo, para que nadie se me ofenda). Mis amigues de otros barrios de la Zona Sur, o de la Zona Oeste, o de La Plata, que todavía están deliberando si se autoperciben o no del Conurbano. También mis amigues de Capital, o quienes vienen de todos esos lugares pero han decidido federalizar su amor, o hasta globalizarlo, y viven hoy a un par de horas en avión de distancia. También a la familia elegida que me hice durante el tiempo que viví en Reino Unido, y que hicieron posible retener un balance positivo de ese capítulo de mi historia.

Que se entienda: mis amigues son como mi familia, pero igual. Me pasa lo propio con los compañeros y compañeras de mis papás, a quienes también adopto y por quienes me dejo adoptar. Hay una relación de intimidad que es sólo propia de ese tipo de vínculo, de una forma inintercambiable, irremplazable, como son imposibles de reemplazar las ausencias de quienes ya me tocó perder también de esta familia. Mis amigues son el medio, pero también el fin. Son el objetivo, pero también la metodología. Son toda la belleza que me puedo merecer, y hasta más. Por cierto, ¿han visto lo bonita que está esa portada de tesis? Se lo debo a Pau, que es una parte fundamental de esa familia elegida, sobre todo porque fue quien más y

mejor me enseñó la importancia de construir una, desde el más profundo sentido de amor y compañía. A Pau, y al resto de mi familia elegida, también les corresponde un agradecimiento especial, escrito entre pucheros.

Pero no vaya a pensarse que no queda nadie de mi casa de los acentos raros. Las excepciones a las reglas son, si se me permite la redundancia, excepcionales. Queda mi viejo, por ejemplo. Queda mi viejo, y queda el arco narrativo de mi viejo. La referencia de cómo resulta posible vivir una vida de reivindicaciones inclaudicables y superar holgadamente el punto a partir del cual uno ya debería haber vendido sus principios, y no hacerlo. Mi viejo es un complot contra el tiempo y las fechas de vencimiento. "Ganaremos cada vez que algún joven lea estas líneas y sepa que no todo se compra ni se vende y sienta ganas de querer cambiar el mundo". Yo fui ese joven. Y fue gracias a mi viejo. A él también mi agradecimiento.

Me queda también mi tío, el hermano de mi vieja, y quien quizás fue cómplice inconsciente de haber escrito esta tesis, hace muchos años. Porque le atribuyo buena parte del inicio a mi vocación científica a él, egresado también de nuestra querida UNQ. Su hijo Lautaro, mi primo, es a su vez la prueba vigente de que también los hubo días buenos, veranos interminables, piezas de vida que no son negociables. Lautaro es hacer del recuerdo una cotidianidad que construye un nuevo recuerdo. El eterno retorno, pero bien.

A mi tío le atribuyo junto con su compañera la autoría de la obra que me motivó a continuar con la escritura en los momentos críticos que toda tesis apareja. Esa obra es Pilar, la nieta de mis abuelos, la sobrina de mi vieja, mi prima. Pilar es un recuerdo del futuro. Es una trampa contra el tiempo, como si mirar al cielo no fuera un retrato del pasado sino una foto de lo que está por venir. Un cielo luminoso, un cielo fantástico, un cielo de estrellas fugaces y de otras más estáticas que dibujan la silueta de cosas hermosas e imposibles. Mi prima es una corazonada, o más bien una certeza, de que nos esperan cosas mejores. Será nuestra responsabilidad reservarnos el pesimismo hasta entonces.

Y quedan también quienes ya no están. Vaya si nos quedan. Tiendo a pensar que la forma en la que quedan es a través de lo que elegimos reproducir de su recuerdo. Sus ideas, sus formas, sus gestos de manera inconsciente. Recordar es una forma

de percibir, y en general es constructiva, porque los árboles caídos también son el bosque. Ya repasé esas ausencias en los párrafos anteriores, pero no quiero dejar de hacer una mención especial a mi vieja. Especial, porque así era ella. Mi vieja es todo lo que entiendo por noción de humanidad. Es la prueba de que, como diría Gata Cattana, "algunos eran buenos / algunos no eran prescindibles". Que de toda nuestra obra como humanidad, una parte se salva. Mi vieja es esa parte que se salva, como lo es también la multitud a la que ella supo salvar. Me incluyo.

Quiero despejar la nube. Inhalo, exhalo. Pienso en otras personas a las que me surge agradecer en este punto. Pienso en mis estudiantes y en todo lo que me enseñaron sobre mi propia tesis. Pienso en Alexandra Elbakyan, y su compromiso por la democratización de la Ciencia que hizo posible desarrollar esta investigación. Pienso en el Club Atlético Boca Juniors, su síntesis popular, a modo de resumen gozoso de mi parte menos amarga. Y en las enormes alegrías que el fútbol también supo darme a través de la camiseta celeste y blanca, y particularmente, de la dorsal con el número 10. Pienso en la banda sonora que me acompañó durante todos estos años, cómplice necesaria de esta tesis. Pienso en algunos libros, pienso en algunas series, pienso en las maravillas naturales que tiene este pedazo recóndito del planeta, y que nadie más tiene.

Pienso en los 30 mil ejemplos de resistencia frente a lo que no debiera ocurrir de nuevo, nunca más. Pienso en pibes jóvenes y mal armados pasando frío en una isla lejana. Y pienso en un montón de madres y abuelas a las que deberíamos agradecer por no permitir que envejezca nuestra memoria, y nuestro sentido de humanidad. No alcanzaría otra tesis para agradecerles a ellas también.

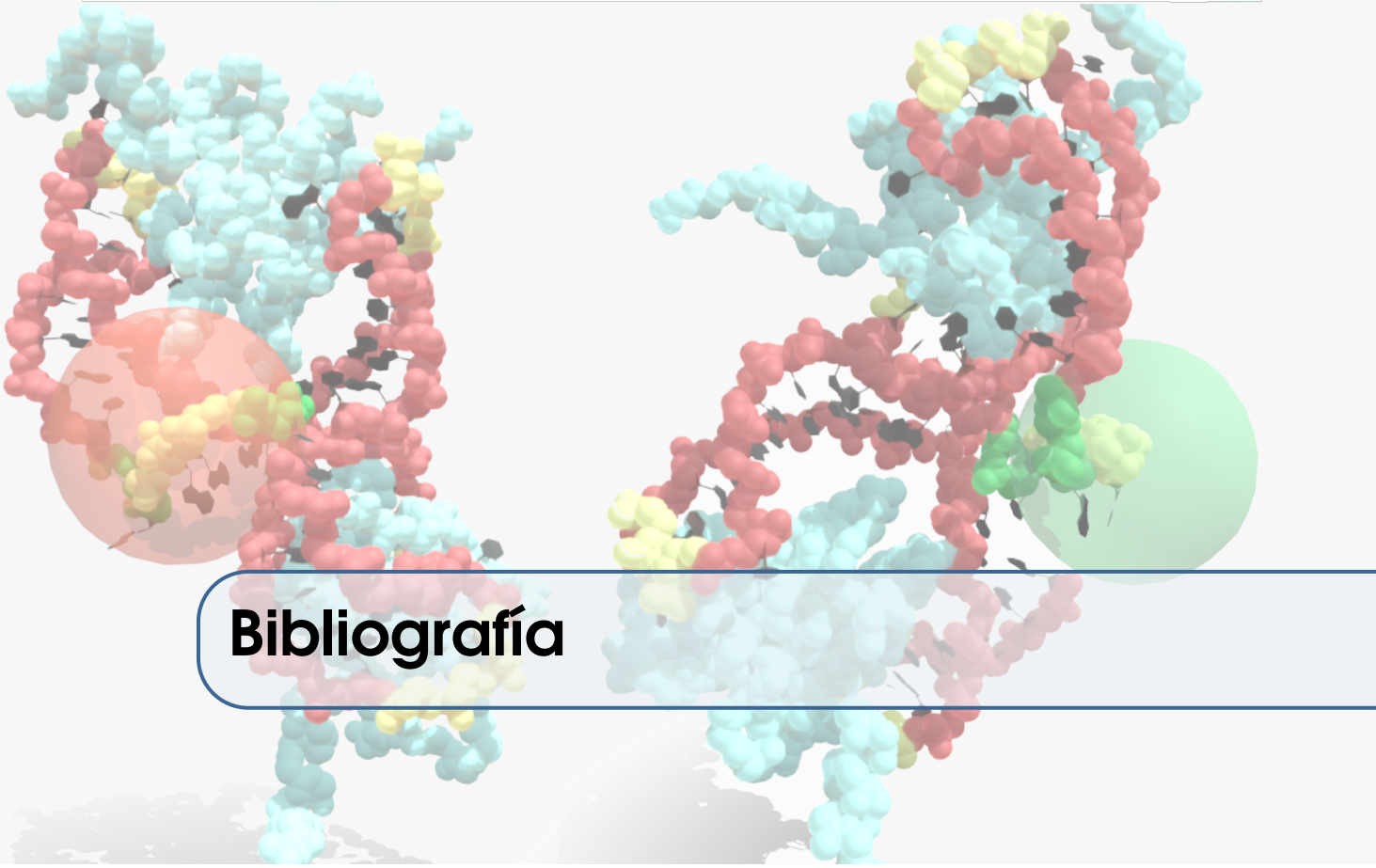
También pienso en lo que dejé para el final a propósito, que es en la gente de mi laboratorio, porque me parece el mejor cierre para esta tesis. Parte de mi familia elegida fue ese laboratorio en el que elegí escribir este pedazo de mi historia. Mis compañeros y compañeras de estos años, y esa humanidad compartida frente a las vicisitudes de la vida doctoral. Mi primera experiencia de paternidad académica fue también en ese laboratorio, con los enormes aprendizajes bidireccionales a los que estoy profundamente agradecido. Y fue también gracias a ese laboratorio que tuve las herramientas para dar mis primeros pasos como docente (qué hermosa la

docencia, ¿no?). Pero sobre todo les estoy agradecido a Juliana y Gustavo. La confianza que me tuvieron desde el primer día no me la tuve yo hasta el último. Vieron algo que a mi me resulta difícil de ver, y que irónicamente puedo verlo sólo en elles.

Quien lea estas líneas puede pensar que no es lo más prolijo inmiscuir lo personal en una relación profesional como la que debieran tener directores y becaries. Y tendría razón. Nunca fuimos gente prolija. No se puede ser prolijo cuando en el medio está la vida. Y la vida que nos tocó durante estos años de investigación no fue sencilla. En esa complejidad es que nos encontramos como equipo, como familia, con momentos de enorme alegría y otros en el que los hornos no estuvieron para bollos. No se sale ileso de un proceso como este, como no se sale ileso de la vida. Esta tesis es un registro de una parte de esa vida, nuestra vida, mi propia vida. Juliana y Gustavo fueron protagonistas de esa historia.

Ojalá el próximo capítulo nos encuentre de nuevo.

Gracias por estos años de Ciencia, de amor y de gente,
Agus



Bibliografía

[1] Agustín Ormazábal, Juliana Palma, and Gustavo Pierdominici-Sottile. *Dynamics and Function of sRNA/mRNAs Under the Scrutiny of Computational Simulation Methods*, pages 207–238. Springer US, 2024. ISBN 9781071635650. doi: 10.1007/978-1-0716-3565-0_12. URL http://dx.doi.org/10.1007/978-1-0716-3565-0_12.

[2] Agustín Ormazábal, Juliana Palma, and Gustavo Pierdominici-Sottile. Molecular Dynamics simulations unveil the basis of the sequential binding of RsmE to the noncoding RNA RsmZ. *The Journal of Physical Chemistry B*, 125(12):3045–3056, 2021. URL https://pubs.acs.org/doi/abs/10.1021/acs.jpcb.0c09770.

[3] Agustín Ormazábal, Gustavo Pierdominici-Sottile, and Juliana Palma. Recognition and binding of RsmE to an AGGAC motif of RsmZ: Insights from Molecular Dynamics simulations. *Journal of Chemical Information and Modeling*, 2022. URL https://pubs.acs.org/doi/10.1021/acs.jcim.2c00037.

[4] Agustín Ormazábal, Matías Sebastián Carletti, Tadeo Enrique Saldaño, Martín Gonzalez Buitron, Julia Marchetti, Nicolas Palopoli, and Alex Bateman. Expanding the repertoire of human tandem repeat RNA-binding proteins. *PLOS ONE*, 18(9):e0290890, September 2023. doi: 10.1371/journal.pone.0290890. URL https://doi.org/10.1371/journal.pone.0290890.

[5] Leland S. Pierson III, Derek W. Wood, and Elizabeth A. Pierson. Homoserine lactonemediated gene regulation in plantassociated bacteria. *Annual Review of Phytopathology*, 36(1):207–225, September 1998. doi: 10.1146/annurev.phyto.36.1.207. URL https://doi.org/10.1146/annurev.phyto.36.1.207.

[6] Richard P Novick and Tom W Muir. Virulence gene regulation by peptides in *Staphylococci* and other Gram-positive bacteria. *Current Opinion in Microbiology*, 2(1):40–45, February 1999. doi: 10.1016/s1369-5274(99)80007-1. URL https://doi.org/10.1016/s1369-5274(99)80007-1.

[7] Leo Eberl. N-acyl homoserinelactone-mediated gene regulation in Gram-negative bacteria. *Systematic and Applied Microbiology*, 22(4):493–506, December 1999. doi: 10.1016/s0723-2020(99)80001-0. URL https://doi.org/10.1016/s0723-2020(99)80001-0.

[8] Christopher A. Vakulskas, Anastasia H. Potts, Paul Babitzke, Brian M. M. Ahmer, and Tony Romeo. Regulation of bacterial virulence by Csr (Rsm) systems. *Microbiology and Molecular Biology Reviews*, 79(2):193–224, June 2015. doi: 10.1128/mmbr.00052-14. URL https://doi.org/10.1128/mmbr.00052-14.

[9] Patricio Martín Sobrero and Claudio Valverde. Comparative genomics and evolutionary analysis of RNA-binding proteins of the CsrA family in the genus P*seudomonas*. *Frontiers in Molecular Biosciences*, 7, July 2020. doi: 10.3389/fmolb.2020.00127. URL https://doi.org/10.3389/fmolb.2020.00127.

[10] T Romeo, M Gong, M Y Liu, and A M Brun-Zinkernagel. Identification and molecular characterization of *csr*A, a pleiotropic gene from E*scherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *Journal of Bacteriology*, 175(15):4744–4755, 1993. doi: 10.1128/jb.175.15.4744-4755.1993. URL https://journals.asm.org/doi/abs/10.1128/jb.175.15.4744-4755.1993.

[11] T Romeo and M Gong. Genetic and physical mapping of the regulatory gene *csr*A on the E*scherichia coli* k-12 chromosome. *Journal of Bacteriology*, 175(17):5740–5741, September 1993. doi: 10.1128/jb.175.17.5740-5741.1993. URL https://doi.org/10.1128/jb.175.17.5740-5741.1993.

[12] Tony Romeo, Christopher A. Vakulskas, and Paul Babitzke. Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. *Environmental Microbiology*, 15(2):313–324, 2013. doi: https://doi.org/10.1111/j.1462-2920.2012.02794.x. URL https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1462-2920.2012.02794.x.

[13] Kai Papenfort and Jörg Vogel. Regulatory RNA in bacterial pathogens. *Cell Host & Microbe*, 8(1):116–127, July 2010. doi: 10.1016/j.chom.2010.06.008. URL https://doi.org/10.1016/j.chom.2010.06.008.

[14] Ann Kathrin Heroven, Katja Böhme, and Petra Dersch. The Csr/Rsm system of Y*ersinia* and related pathogens: a post-transcriptional strategy for managing virulence. *RNA biology*, 9(4):379-391, April 2012. ISSN 1547-6286. doi: 10.4161/rna.19333. URL https://doi.org/10.4161/rna.19333.

[15] Tony Romeo. Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Molecular Microbiology*, 29(6):1321–1330, 1998. doi: https://doi.org/10.1046/j.1365-2958.1998.01021.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.1998.01021.x.

[16] Charlotte Michaux, Nicolas Verneuil, Axel Hartke, and Jean-Christophe Giard. Physiological roles of small RNA molecules. *Microbiology*, 160(6):1007–1019, 2014. ISSN 1465-2080. doi: https://doi.org/10.1099/mic.0.076208-0. URL https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.076208-0.

[17] Alejandro Toledo-Arana, Francis Repoila, and Pascale Cossart. Small noncoding RNAs controlling pathogenesis. *Current Opinion in Microbiology*, 10(2):182–188, 2007. ISSN 1369-5274. doi: https://doi.org/10.1016/j.mib.2007.03.004. URL https://www.sciencedirect.com/science/article/pii/S1369527407000215. Cell regulation (RNA special issue).

[18] Debra White, Mark E. Hart, and Tony Romeo. Phylogenetic distribution of the global regulatory gene *csr*A among eubacteria. *Gene*, 182(1-2):221–223, December 1996. doi: 10.1016/s0378-1119(96)00547-1. URL https://doi.org/10.1016/s0378-1119(96)00547-1.

[19] Y Cui, A Chatterjee, Y Liu, C K Dumenyo, and A K Chatterjee. Identification of a global repressor gene, *rsma*, of E*rwinia carotovora* subsp. *carotovora* that controls extracellular enzymes, n-(3-oxohexanoyl)-l-homoserine lactone, and pathogenicity in soft-rotting E*rwinia* spp. *Journal of Bacteriology*, 177(17): 5108–5115, September 1995. doi: 10.1128/jb.177.17.5108-5115.1995. URL https://doi.org/10.1128/jb.177.17.5108-5115.1995.

[20] Valverde Claudio et al. A repeated GGA motif is critical for the activity and stability of the riboregulator RsmY of P*seudomonas fluorescens*. *RNA: Structure, Metabolism and Catalysis*, 279(24):25066–25074, 2004. doi: doi:10.1074/jbc.M401870200. URL https://www.jbc.org/article/S0021-9258(20)66402-X/fulltext.

[21] Adrianne N. Edwards, Laura M. Patterson-Fortin, Christopher A. Vakulskas, Jeffrey W. Mercante, Katarzyna Potrykus, Daniel Vinella, Martha I. Camacho, Joshua A. Fields, Stuart A. Thompson, Dimitris Georgellis, Michael Cashel, Paul Babitzke, and Tony Romeo. Circuitry linking the Csr and stringent response global regulatory systems. *Molecular Microbiology*, 80(6):1561–1580, May 2011. doi: 10.1111/j.1365-2958.2011.07663.x. URL https://doi.org/10.1111/j.1365-2958.2011.07663.x.

[22] A.H. Potts, C.A. Vakulskas, A. Pannuri, and et al. Global role of the bacterial post-transcriptional regulator CsrA revealed by integrated transcriptomics. *Nat Commun*, 8:1596, 2017. doi: https://doi.org/10.1038/s41467-017-01613-1. URL https://doi.org/10.4161/rna.7.5.12687.

[23] Helen Yakhnin, Carol S. Baker, Igor Berezin, Michael A. Evangelista, Alisa Rassin, Tony Romeo, and Paul Babitzke. CsrA represses translation of *sdi*A, which encodes the -acylhomoserine- scp -lactone receptor of E*scherichia coli*, by binding exclusively within the coding region of *sdia* mRNA. *Journal of Bacteriology*, 193(22):6162–6170, November 2011. doi: 10.1128/jb.05975-11. URL https://doi.org/10.1128/jb.05975-11.

[24] M Y Liu, H Yang, and T Romeo. The product of the pleiotropic E*scherichia coli* gene *csr*A modulates glycogen biosynthesis via effects on mRNA stability. *Journal of Bacteriology*, 177(10):2663–2672, May 1995. doi: 10.1128/jb.177.10.2663-2672.1995. URL https://doi.org/10.1128/jb.177.10.2663-2672.1995.

[25] Carol S. Baker, Igor Morozov, Kazushi Suzuki, Tony Romeo, and Paul Babitzke. CsrA regulates glycogen biosynthesis by preventing translation of glgC in E*scherichia coli*. *Molecular Microbiology*, 44(6):1599–1610, 2002. doi: https://doi.org/10.1046/j.1365-2958.2002.02982.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.2002.02982.x.

[26] Mario Schubert, Karine Lapouge, Olivier Duss, Florian C Oberstrass, Ilian Jelesarov, Dieter Haas, and Frédéric H-T Allain. Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. *Nature Structural Molecular Biology*, 14(9):807–813, August 2007. doi: 10.1038/nsmb1285. URL https://doi.org/10.1038/nsmb1285.

[27] A. Pannuri, H. Yakhnin, C. A. Vakulskas, A. N. Edwards, P. Babitzke, and T. Romeo. Translational repression of NhaR, a novel pathway for multi-tier regulation of biofilm circuitry by CsrA. *Journal of*

Bacteriology, 194(1):79–89, October 2011. doi: 10.1128/jb.06209-11. URL https://doi.org/10.1128/jb.06209-11.

[28] Yasuhiko Irie, Melissa Starkey, Adrianne N. Edwards, Daniel J. Wozniak, Tony Romeo, and Matthew R. Parsek. P*seudomonas aeruginosa* biofilm matrix polysaccharide psl is regulated transcriptionally by RpoS and post-transcriptionally by RsmA. *Molecular Microbiology*, pages no–no, September 2010. doi: 10.1111/j.1365-2958.2010.07320.x. URL https://doi.org/10.1111/j.1365-2958.2010.07320.x.

[29] Xin Wang, Ashok K. Dubey, Kazushi Suzuki, Carol S. Baker, Paul Babitzke, and Tony Romeo. CsrA post-transcriptionally represses *pga*ABCD, responsible for synthesis of a biofilm polysaccharide adhesin of E*scherichia coli*. *Molecular Microbiology*, 56(6):1648–1663, April 2005. doi: 10.1111/j.1365-2958.2005.04648.x. URL https://doi.org/10.1111/j.1365-2958.2005.04648.x.

[30] Laura M Patterson-Fortin, C. Vakulskas, H. Yakhnin, P. Babitzke, and T. Romeo. Dual posttranscriptional regulation via a cofactor-responsive mRNA leader. *Journal of molecular biology*, 425 19:3662–77, 2013. doi: 10.1016/j.jmb.2012.12.010. URL https://www.sciencedirect.com/science/article/abs/pii/S0022283612009436?via%3Dihub.

[31] Bangdong L. Wei, Anne-Marie Brun-Zinkernagel, Jerry W. Simecka, Birgit M. Prüß, Paul Babitzke, and Tony Romeo. Positive regulation of motility and *flh*DC expression by the RNA-binding protein CsrA of E*scherichia coli*. *Molecular Microbiology*, 40(1):245–256, 2001. doi: https://doi.org/10.1046/j.1365-2958.2001.02380.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.2001.02380.x.

[32] Alexander V. Yakhnin, Carol S. Baker, Christopher A. Vakulskas, Helen Yakhnin, Igor Berezin, Tony Romeo, and Paul Babitzke. CsrA activates *flh*DC expression by protecting *flh*DC mRNA from RNAse E-mediated cleavage. *Molecular Microbiology*, 87(4):851–866, 2013. doi: https://doi.org/10.1111/mmi.12136. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mmi.12136.

[33] T. Romeo. Post-transcriptional regulation of bacterial carbohydrate metabolism: evidence that the gene product CsrA is a global mRNA decay factor. *Research in Microbiology*, 147(6-7):505–512, July 1996. doi: 10.1016/0923-2508(96)84004-6. URL https://doi.org/10.1016/0923-2508(96)84004-6.

[34] Pablo Gutiérrez, Yan Li, Michael J. Osborne, Ekaterina Pomerantseva, Qian Liu, and Kalle Gehring. Solution structure of the carbon storage regulator protein CsrA from E*scherichia coli*. *Journal of Bacteriology*, 187(10):3496–3501, 2005. doi: 10.1128/JB.187.10.3496-3501.2005. URL https://journals.asm.org/doi/abs/10.1128/JB.187.10.3496-3501.2005.

[35] Jeffrey Mercante, Kazushi Suzuki, Xiaodong Cheng, Paul Babitzke, and Tony Romeo. Comprehensive Alanine-scanning mutagenesis of E*scherichia coli* CsrA defines two subdomains of critical functional importance. *Journal of Biological Chemistry*, 281(42):31832–31842, August 2006. doi: 10.1074/jbc.m606057200. URL https://doi.org/10.1074/jbc.m606057200.

[36] Stephan Heeb, Sarah A. Kuehne, Mark Bycroft, Sorana Crivii, Mark D. Allen, Dieter Haas, Miguel Cámara, and Paul Williams. Functional analysis of the post-transcriptional regulator RsmA reveals a novel RNA-binding site. *Journal of Molecular Biology*, 355(5):1026–1036, 2006. ISSN 0022-2836. doi: https://doi.org/10.1016/j.jmb.2005.11.045. URL https://www.sciencedirect.com/science/article/pii/S0022283605014245.

[37] Dubey AK, Baker CS, Romeo T, and Babitzke P. SRNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA*, 11(10):1579-1587, 2005. doi: 10.1261/rna.2990205. URL https://rnajournal.cshlp.org/content/11/10/1579.

[38] Seshagirirao Gudapaty, Kazushi Suzuki, Xin Wang, Paul Babitzke, and Tony Romeo. Regulatory interactions of Csr components: the RNA binding protein CsrA activates *csr*B transcription in E*scherichia coli*. *Journal of Bacteriology*, 183(20):6017–6027, October 2001. doi: 10.1128/jb.183.20.6017-6027.2001. URL https://doi.org/10.1128/jb.183.20.6017-6027.2001.

[39] Sampriti Mukherjee, Helen Yakhnin, Dave Kysela, Josh Sokoloski, Paul Babitzke, and Daniel B. Kearns. CsrA - FliW interaction governs flagellin homeostasis and a checkpoint on flagellar morphogenesis in B*acillus subtilis*. *Molecular Microbiology*, 82(2):447–461, 2011. doi: https://doi.org/10.1111/j.1365-2958.2011.07822.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2011.07822.x.

[40] Sampriti Mukherjee, Paul Babitzke, and Daniel B. Kearns. FliW and FliS function independently to control cytoplasmic flagellin levels in B*acillus subtilis*. *Journal of Bacteriology*, 195(2):297–306, January 2013. doi: 10.1128/jb.01654-12. URL https://doi.org/10.1128/jb.01654-12.

[41] Katarzyna A. Radomska, Soledad R. Ordoñez, Marc M. S. M. Wösten, Jaap A. Wagenaar, and Jos P. M. van Putten. Feedback control of C*ampylobacter jejuni* flagellin levels through reciprocal binding of FliW to flagellin and the global regulator CsrA. *Molecular Microbiology*, 102(2):207–220, July 2016. doi: 10.1111/mmi.13455. URL https://doi.org/10.1111/mmi.13455.

[42] Sampriti Mukherjee, Reid T. Oshiro, Helen Yakhnin, Paul Babitzke, and Daniel B. Kearns. FliW antagonizes CsrA RNA binding by a noncompetitive allosteric mechanism. *Proceedings of the National Academy of Sciences*, 113(35):9870–9875, 2016. ISSN 0027-8424. doi: 10.1073/pnas.1602455113. URL https://www.pnas.org/content/113/35/9870.

[43] Florian Altegoer, Stefan A. Rensing, and Gert Bange. Structural basis for the CsrA-dependent modulation of translation initiation by an ancient regulatory protein. *Proceedings of the National Academy of Sciences*, 113(36):10168–10173, 2016. ISSN 0027-8424. doi: 10.1073/pnas.1602425113. URL https://www.pnas.org/content/113/36/10168.

[44] Ashok K. Dubey, Carol S. Baker, Kazushi Suzuki, A. Daniel Jones, Pallavi Pandit, Tony Romeo, and Paul Babitzke. CsrA regulates translation of the E*scherichia coli* carbon starvation gene, *csr*A, by blocking ribosome access to the *cst*A transcript. *Journal of Bacteriology*, 185(15):4450–4460, August 2003. doi: 10.1128/jb.185.15.4450-4460.2003. URL https://doi.org/10.1128/jb.185.15.4450-4460.2003.

[45] Carol S. Baker, Lél A. Eöry, Helen Yakhnin, Jeffrey Mercante, Tony Romeo, and Paul Babitzke. CsrA inhibits translation initiation of E*scherichia coli hfq* by binding to a single site overlapping the Shine Delgarno sequence. *Journal of Bacteriology*, 189(15):5472–5481, 2007. doi: 10.1128/JB.00529-07. URL https://journals.asm.org/doi/abs/10.1128/JB.00529-07.

[46] Helen Yakhnin, Alexander V. Yakhnin, Carol S. Baker, Elena Sineva, Igor Berezin, Tony Romeo, and Paul Babitzke. Complex regulation of the global regulatory gene *csr*A: CsrA-mediated translational repression, transcription from five promoters by e$\sigma^7$0 and eσ(s), and indirect transcriptional activation by CsrA. *Molecular Microbiology*, 81(3):689–704, June 2011. doi: 10.1111/j.1365-2958.2011.07723.x. URL https://doi.org/10.1111/j.1365-2958.2011.07723.x.

[47] Mu Ya et al. Liu. The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in E*scherichia coli*. *JBC*, 272(78):17502–17510, 1997. doi: 10.1074/jbc.272.28.17502. URL https://www.jbc.org/article/S0021-9258(18)39224-X/fulltext.

[48] Tony Romeo, Paul Babitzke, Gisela Storz, and Kai Papenfort. Global regulation by CsrA and its RNA antagonists. *Microbiology Spectrum*, 6(2):6.2.05, 2018. doi: 10.1128/microbiolspec.RWR-0009-2017. URL https://journals.asm.org/doi/abs/10.1128/microbiolspec.RWR-0009-2017.

[49] Li Mei, Sanger Xu, Peng Lu, Haiping Lin, Yanbin Guo, and Yongjun Wang. Csrb, a noncoding regulatory rna, is required for bara-dependent expression of biocontrol traits in rahnella aquatilis hx2. *PLOS ONE*, 12(11):e0187492, November 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0187492. URL http://dx.doi.org/10.1371/journal.pone.0187492.

[50] Thomas Weilbacher, Kazushi Suzuki, Ashok K. Dubey, Xin Wang, Seshigirao Gudapaty, Igor Morozov, Carol S. Baker, Dimitris Georgellis, Paul Babitzke, and Tony Romeo. A novel sRNA component of the carbon storage regulatory system of E*scherichia coli*. *Molecular Microbiology*, 48(3):657–670, 2003. doi: https://doi.org/10.1046/j.1365-2958.2003.03459.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.2003.03459.x.

[51] Mikkel Girke Jørgensen, Maureen K. Thomason, Johannes Havelund, Poul Valentin-Hansen, and Gisela Storz. Dual function of the McaS small RNA in controlling biofilm formation. *Genes & Development*, 27(10):1132–1145, May 2013. doi: 10.1101/gad.214734.113. URL https://doi.org/10.1101/gad.214734.113.

[52] Luary C. Martínez and Viveka Vadyvaloo. Mechanisms of post-transcriptional gene regulation in bacterial biofilms. *Frontiers in Cellular and Infection Microbiology*, 4, March 2014. doi: 10.3389/fcimb.2014.00038. URL https://doi.org/10.3389/fcimb.2014.00038.

[53] Kristina Jonas and Öjar Melefors. The E*scherichia coli* CsrB and CsrC small RNAs are strongly induced during growth in nutrient-poor medium. *FEMS Microbiology Letters*, 297(1):80–86, August 2009. doi: 10.1111/j.1574-6968.2009.01661.x. URL https://doi.org/10.1111/j.1574-6968.2009.01661.x.

[54] Maureen K. Thomason, Fanette Fontaine, Nicholas De Lay, and Gisela Storz. A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in E*scherichia coli*. *Molecular Microbiology*, 84(1):17–35, January 2012. doi: 10.1111/j.1365-2958.2012.07965.x. URL https://doi.org/10.1111/j.1365-2958.2012.07965.x.

[55] E Gerhart H Wagner. Kill the messenger: bacterial antisense RNA promotes mRNA decay. *Nature Structural & Molecular Biology*, 16(8):804–806, August 2009. doi: 10.1038/nsmb0809-804. URL https://doi.org/10.1038/nsmb0809-804.

[56] Esquerré T, Bouvier M, Turlan C, Carpousis AJ, Girbal L, and Cocaign-Bousquet M. The Csr system regulates genome-wide mRNA stability and transcription and thus gene expression in E*scherichia coli*. *Scientific Reports*, 6(1), April 2016. doi: 10.1038/srep25057. URL https://doi.org/10.1038/srep25057.

[57] Paul Babitzke and Tony Romeo. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Current Opinion in Microbiology*, 10(2):156–163, 2007. ISSN 1369-5274. doi: https://doi.org/10.1016/j.mib.2007.03.007. URL https://www.sciencedirect.com/science/article/pii/S1369527407000240. Cell regulation (RNA special issue).

[58] Johan Timmermans and Laurence Van Melderen. Post-transcriptional global regulation by CsrA in bacteria. *Cellular and Molecular Life Sciences*, 67(17):2897–2908, May 2010. doi: 10.1007/s00018-010-0381-z. URL https://doi.org/10.1007/s00018-010-0381-z.

[59] Luary C. Martínez, Helen Yakhnin, Martha I. Camacho, Dimitris Georgellis, Paul Babitzke, José L. Puente, and Víctor H. Bustamante. Integration of a complex regulatory cascade involving the SirA/BarA and Csr global regulatory systems that controls expression of the S*almonella* SPI-1 and SPI-2 virulence regulons through HilD. *Molecular Microbiology*, 80(6):1637–1656, 2011. doi: https://doi.org/10.1111/j.1365-2958.2011.07674.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2011.07674.x.

[60] Van Melderen L Timmermans J. Regulation by small RNAs in bacteria: Expanding frontiers. *Molecular Cell*, 6(43):880âĂŞ891, 2011. doi: 10.1016/j.molcel.2011.08.022. URL https://www.cell.com/molecular-cell/fulltext/S1097-2765(11)00643-5?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS1097276511006435%3Fshowall%3Dtrue.

[61] Jonas Gripenland, Sakura Netterling, Edmund Loh, Teresa Tiensuu, Alejandro Toledo-Arana, and Jörgen Johansson. RNAs: regulators of bacterial virulence. *Nature Reviews Microbiology*, 8(12):857–866, November 2010. doi: 10.1038/nrmicro2457. URL https://doi.org/10.1038/nrmicro2457.

[62] Reik W Ponting CP, Oliver PL. Evolution and functions of long noncoding RNAs. *Nature*, 4 (136):629–641, 2009. doi: 10.1016/j.cell.2009.02.006. URL https://www.cell.com/cell/fulltext/S0092-8674(09)00142-1?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867409001421%3Fshowall%3Dtrue.

[63] Elisabeth Kay, Christophe Dubuis, and Dieter Haas. Three small RNAs jointly ensure secondary metabolism and biocontrol in P*seudomonas fluorescens* CHA0. *Proceedings of the National Academy of Sciences*, 102(47):17136–17141, November 2005. doi: 10.1073/pnas.0505673102. URL https://doi.org/10.1073/pnas.0505673102.

[64] Carrie Selin, Jerrylynn Manuel, W.G. Dilantha Fernando, and Teresa de Kievit. Expression of the P*seudomonas chlororaphis* strain PA23 Rsm system is under control of GacA, RpoS, PsrA, quorum sensing and the stringent response. *Biological Control*, 69:24–33, 2014. ISSN 1049-9644. doi: https://doi.org/10.1016/j.biocontrol.2013.10.015. URL https://www.sciencedirect.com/science/article/pii/S1049964413002454.

[65] Lamia Harfouche, Feth el Zahar Haichar, and Wafa Achouak. Small regulatory RNAs and the fine-tuning of plant - bacteria interactions. *New Phytologist*, 206(1):98–106, 2015. doi: https://doi.org/10.1111/nph.13195. URL https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.13195.

[66] Patricio Martín Sobrero, Andrés Muzlera, Julieta Frescura, Edgardo Jofré, and Claudio Valverde. A matter of hierarchy: activation of orfamide production by the post-transcriptional Gac-Rsm cascade of P*seudomonas protegens* CHA0 through expression upregulation of the two dedicated transcriptional regulators. *Environmental Microbiology Reports*, 9(5):599–611, 2017. doi: https://doi.org/10.1111/1758-2229.12566. URL https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1758-2229.12566.

[67] Anja Brencic, Kirsty A. McFarland, Heather R. McManus, Sandra Castang, Ilaria Mogno, Simon L. Dove, and Stephen Lory. The GacS/GacA signal transduction system of *Pseudomonas aeruginosa* acts exclusively through its control over the transcription of the RsmY and RsmZ regulatory small RNAs. *Molecular Microbiology*, 73(3):434–445, 2009. doi: https://doi.org/10.1111/j.1365-2958.2009.06782.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2009.06782.x.

[68] M.C. Alvarez Crespo and C Valverde. A single mutation in the oprF mRNA leader confers strict translational control by the Gac/Rsm system in *Pseudomonas fluorescens* CHA0. *Curr Microbiol*, 58: 182-188, 2009. doi: https://doi.org/10.1007/s00284-008-9306-6. URL https://link.springer.com/article/10.1007/s00284-008-9306-6.

[69] Karine Lapouge, Mario Schubert, Frédéric H.-T. Allain, and Dieter Haas. Gac/Rsm signal transduction pathway of γ-proteobacteria: from RNA recognition to regulation of social behaviour. *Molecular Microbiology*, 67(2):241–253, 2008. doi: https://doi.org/10.1111/j.1365-2958.2007.06042.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2007.06042.x.

[70] David Lalaouna, Sylvain Fochesato, Lisa Sanchez, Philippe Schmitt-Kopplin, Dieter Haas, Thierry Heulin, and Wafa Achouak. Phenotypic switching in *Pseudomonas brassicacearum* involves GacS- and GacA-dependent Rsm small RNAs. *Applied and Environmental Microbiology*, 78(6):1658–1665, 2012. doi: 10.1128/AEM.06769-11. URL https://journals.asm.org/doi/abs/10.1128/AEM.06769-11.

[71] Anja Brencic and Stephen Lory. Determination of the regulon and identification of novel mRNA targets of *Pseudomonas aeruginosa* RsmA. *Molecular Microbiology*, 72(3):612–632, 2009. doi: https://doi.org/10.1111/j.1365-2958.2009.06670.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2009.06670.x.

[72] Olivier Duss, Erich Michel, Nana Diarra dit Konté, Mario Schubert, and Frédéric H.-T. Allain. Molecular basis for the wide range of affinity found in Csr/Rsm protein - RNA recognition. *Nucleic Acids Research*, 42(8):5332–5346, 02 2014. ISSN 0305-1048. doi: 10.1093/nar/gku141. URL https://doi.org/10.1093/nar/gku141.

[73] Olivier Duss, Erich Michel, Maxim Yulikov, Mario Schubert, Gunnar Jeschke, and Frédéric H.-T. Allain. Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature*, 509(7502): 588–592, May 2014. doi: 10.1038/nature13271. URL https://doi.org/10.1038/nature13271.

[74] Xiang-Jun Lu, Harmen J. Bussemaker, and Wilma K. Olson. DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Research*, page gkv716, July 2015. doi: 10.1093/nar/gkv716. URL https://doi.org/10.1093/nar/gkv716.

[75] Kévin Darty, Alain Denise, and Yann Ponty. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, 25(15):1974–1975, April 2009. doi: 10.1093/bioinformatics/btp250. URL https://doi.org/10.1093/bioinformatics/btp250.

[76] Kayley H. Janssen, Manisha R. Diaz, Cindy J. Gode, Matthew C. Wolfgang, and Timothy L. Yahr. RsmV, a small noncoding regulatory RNA in *Pseudomonas aeruginosa* that sequesters RsmA and RsmF from target mRNAs. *Journal of Bacteriology*, 200(16), August 2018. doi: 10.1128/jb.00277-18. URL https://doi.org/10.1128/jb.00277-18.

[77] C.L. Miller, M. Romero, S.L.R. Karna, and et al. RsmW, *Pseudomonas aeruginosa* small non-coding RsmA-binding RNA upregulated in biofilm versus planktonic growth conditions. *BMC Microbiol*,

16(155), 2016. doi: 10.1186/s12866-016-0771-y. URL https://bmcmicrobiol.biomedcentral.com/articles/10.1186/s12866-016-0771-y.

[78] Xinyu Jia, Zhiling Pan, Yang Yuan, Bingnan Luo, Yongbo Luo, Sunandan Mukherjee, Guowen Jia, Liu Liu, Xiaobin Ling, Xiting Yang, Zhichao Miao, Xiawei Wei, Janusz M. Bujnicki, Kelei Zhao, and Zhaoming Su. Structural basis of sRNA RsmZ regulation of P*seudomonas aeruginosa virulence*. *Cell Research*, 33(4):328–330, February 2023. doi: 10.1038/s41422-023-00786-3. URL https://doi.org/10.1038/s41422-023-00786-3.

[79] Karine Lapouge, Remo Perozzo, Justyna Iwaszkiewicz, Claire Bertelli, Vincent Zoete, Olivier Michielin, Leonardo Scapozza, and Dieter Haas. RNA pentaloop structures as effective targets of regulators belonging to the RsmA/CsrA protein family. *RNA Biology*, 10(6):1030–1041, 2013. doi: 10.4161/rna.24771. URL https://doi.org/10.4161/rna.24771.

[80] Xiaodong Ren, Rui Zeng, Micky Tortorella, Jinming Wang, and Changwei Wang. Structural insight into inhibition of CsrA-RNA interaction revealed by docking, molecular dynamics and free energy calculations. *Scientific Reports*, 7(1), November 2017. doi: 10.1038/s41598-017-14916-6. URL https://doi.org/10.1038/s41598-017-14916-6.

[81] Caroline Blumer, Stephan Heeb, Gabriella Pessi, and Dieter Haas. Global GacA-steered control of cyanide and exoprotease production in P*seudomonas fluorescens* involves specific ribosome binding sites. *Proceedings of the National Academy of Sciences*, 96(24):14073–14078, 1999. ISSN 0027-8424. doi: 10.1073/pnas.96.24.14073. URL https://www.pnas.org/content/96/24/14073.

[82] Laville J, Voisard C, Keel C, Maurhofer M, Défago G, and Haas D. Global control in P*seudomonas fluorescens* mediating antibiotic synthesis and suppression of black root rot of tobacco. *Proc Natl Acad Sci*, 89(5), 1992. doi: 10.1073/pnas.89.5.1562. URL https://www.pnas.org/doi/abs/10.1073/pnas.89.5.1562.

[83] Reimmann C, Beyeler M, Latifi A, Winteler H, Foglino M, Lazdunski A, and Haas D. The global activator GacA of P*seudomonas aeruginosa* PAO positively controls the production of the autoinducer n-butyryl-homoserine lactone and the formation of the virulence factors pyocyanin, cyanide, and lipase. *Mol Microbiol*, 24(2), 1997. doi: 10.1046/j.1365-2958.1997.3291701.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.1997.3291701.x.

[84] Anju Sehrawat, Satyavir S. Sindhu, and Bernard R. Glick. Hydrogen cyanide production by soil bacteria: Biological control of pests and promotion of plant growth in sustainable agriculture. *Pedosphere*, 32(1): 15–38, February 2022. doi: 10.1016/s1002-0160(21)60058-9. URL https://doi.org/10.1016/s1002-0160(21)60058-9.

[85] Haiyun Ma, Xinyu Jia, Kaiming Zhang, and Zhaoming Su. Cryo-EM advances in RNA structure determination. *Signal Transduction and Targeted Therapy*, 7(1), February 2022. doi: 10.1038/s41392-022-00916-0. URL https://doi.org/10.1038/s41392-022-00916-0.

[86] Jiří Šponer, Giovanni Bussi, Miroslav Krepl, Pavel Banáš, Sandro Bottaro, Richard A. Cunha, Alejandro Gil-Ley, Giovanni Pinamonti, Simón Poblete, Petr Jurečka, Nils G. Walter, and Michal Otyepka. RNA structural dynamics as captured by molecular simulations: A comprehensive overview. *Chemical Reviews*, 118(8):4177–4338, 2018. doi: 10.1021/acs.chemrev.7b00427. URL https://pubs.acs.org/doi/10.1021/acs.chemrev.7b00427.

[87] Peter Tompa and Peter Csermely. The role of structural disorder in the function of RNA and protein chaperones. *The FASEB Journal*, 18(11):1169–1175, August 2004. doi: 10.1096/fj.04-1584rev. URL https://doi.org/10.1096/fj.04-1584rev.

[88] Brian FC Clark. The crystal structure of tRNA. *Journal of biosciences*, 31(4):453–457, 2006. URL https://link.springer.com/article/10.1007/BF02705184.

[89] Robert W. Holley. Structure of an alanine transfer ribonucleic acid. *JAMA: The Journal of the American Medical Association*, 194(8):868, November 1965. doi: 10.1001/jama.1965.03090210032009. URL https://doi.org/10.1001/jama.1965.03090210032009.

[90] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*, 181(4610): 662–666, March 1958. doi: 10.1038/181662a0. URL https://doi.org/10.1038/181662a0.

[91] Jiří Šponer, Miroslav Krepl, Pavel Banáš, Petra KÃijhrová, Marie Zgarbová, Petra Jurečka, Marek Havrila, and Michal Otyepka. How to understand atomistic Molecular Dynamics simulations of RNA and protein–RNA complexes? *Wiley Interdisciplinary Reviews: RNA*, 8(3):e1405, 2017. doi: 10.1002/wrna.1405. URL https://wires.onlinelibrary.wiley.com/doi/10.1002/wrna.1405.

[92] Kevin V. Morris and John S. Mattick. The rise of regulatory RNA. *Nature Reviews Genetics*, 15(6): 423–437, April 2014. doi: 10.1038/nrg3722. URL https://doi.org/10.1038/nrg3722.

[93] Kelly Kruger, Paula J. Grabowski, Arthur J. Zaug, Julie Sands, Daniel E. Gottschling, and Thomas R. Cech. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell*, 31(1):147–157, November 1982. doi: 10.1016/0092-8674(82)90414-7. URL https://doi.org/10.1016/0092-8674(82)90414-7.

[94] Gail Mitchell Emmilson, Shingo Namakura, Adam Roth, and Ronald D Breaker. Ribozyme speed limits. *RNA*, 9(8):907–918, July 2003. doi: 10.1261/rna.5680603. URL https://doi.org/10.1261/rna.5680603.

[95] F.H.C. Crick. The origin of the genetic code. *Journal of Molecular Biology*, 38(3):367–379, December 1968. doi: 10.1016/0022-2836(68)90392-6. URL https://doi.org/10.1016/0022-2836(68)90392-6.

[96] L.E. Orgel. Evolution of the genetic apparatus. *Journal of Molecular Biology*, 38(3):381–393, December 1968. doi: 10.1016/0022-2836(68)90393-8. URL https://doi.org/10.1016/0022-2836(68)90393-8.

[97] C R Woese, D H Dugre, W C Saxinger, and S A Dugre. The molecular basis for the genetic code. *Proceedings of the National Academy of Sciences*, 55(4):966–974, April 1966. doi: 10.1073/pnas.55.4.966. URL https://doi.org/10.1073/pnas.55.4.966.

[98] Biliang Zhang and Thomas R. Cech. Peptidyl-transferase ribozymes: trans reactions, structural characterization and ribosomal RNA-like features. *Chemistry &amp Biology*, 5(10):539–553, October 1998. doi: 10.1016/s1074-5521(98)90113-2. URL https://doi.org/10.1016/s1074-5521(98)90113-2.

[99] David H. Mathews, Jeffrey Sabina, Michael Zuker, and Douglas H. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology*, 288(5):911–940, May 1999. doi: 10.1006/jmbi.1999.2700. URL https://doi.org/10.1006/jmbi.1999.2700.

[100] E. Kierzek. Nearest neighbor parameters for Watson-Crick complementary heteroduplexes formed between 2'-O-methyl RNA and RNA oligonucleotides. *Nucleic Acids Research*, 34(13):3609–3614, July 2006. doi: 10.1093/nar/gkl232. URL https://doi.org/10.1093/nar/gkl232.

[101] D.H. Turner, N. Sugimoto, J.A. Jaeger, C.E. Longfellow, S.M. Freier, and R. Kierzek. Improved parameters for prediction of RNA structure. *Cold Spring Harbor Symposia on Quantitative Biology*, 52 (0):123–133, January 1987. doi: 10.1101/sqb.1987.052.01.017. URL https://doi.org/10.1101/sqb.1987.052.01.017.

[102] K. Hoogsteen. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica*, 16(9):907–916, September 1963. doi: 10.1107/s0365110x63002437. URL https://doi.org/10.1107/s0365110x63002437.

[103] A. Mokdad. Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Research*, 34(5):1326–1341, March 2006. doi: 10.1093/nar/gkl025. URL https://doi.org/10.1093/nar/gkl025.

[104] Jiří Šponer, Judit E. Šponer, Anton I. Petrov, and Neocles B. Leontis. Quantum chemical studies of nucleic acids: Can we construct a bridge to the RNA structural biology and bioinformatics communities? *The Journal of Physical Chemistry B*, 114(48):15723–15741, November 2010. doi: 10.1021/jp104361m. URL https://doi.org/10.1021/jp104361m.

[105] Michael Sarver, Craig L. Zirbel, Jesse Stombaugh, Ali Mokdad, and Neocles B. Leontis. FR3d: finding local and composite recurrent structural motifs in RNA 3D structures. *Journal of Mathematical Biology*, 56(1-2):215–252, August 2007. doi: 10.1007/s00285-007-0110-x. URL https://doi.org/10.1007/s00285-007-0110-x.

[106] Anton I. Petrov, Craig L. Zirbel, and Neocles B. Leontis. Automated classification of RNA 3D motifs and the RNA 3D motif atlas. *RNA*, 19(10):1327–1340, August 2013. doi: 10.1261/rna.039438.113. URL https://doi.org/10.1261/rna.039438.113.

[107] Neocles B Leontis and Eric Westhof. Analysis of RNA motifs. *Current Opinion in Structural Biology*, 13(3):300–308, June 2003. doi: 10.1016/s0959-440x(03)00076-9. URL https://doi.org/10.1016/s0959-440x(03)00076-9.

[108] A. Lescoute. Recurrent structural RNA motifs, isostericity matrices and sequence alignments. *Nucleic Acids Research*, 33(8):2395–2409, April 2005. doi: 10.1093/nar/gki535. URL https://doi.org/10.1093/nar/gki535.

[109] Corinna Theis, Craig L. Zirbel, Christian Höner zu Siederdissen, Christian Anthon, Ivo L. Hofacker, Henrik Nielsen, and Jan Gorodkin. RNA 3D modules in genome-wide predictions of RNA 2D structure. *PLOS ONE*, 10(10):e0139900, October 2015. doi: 10.1371/journal.pone.0139900. URL https://doi.org/10.1371/journal.pone.0139900.

[110] Lorena G. Parlea, Blake A. Sweeney, Maryam Hosseini-Asanjan, Craig L. Zirbel, and Neocles B. Leontis. The RNA 3D motif atlas: Computational methods for extraction, organization and evaluation of RNA motifs. *Methods*, 103:99–119, July 2016. doi: 10.1016/j.ymeth.2016.04.025. URL https://doi.org/10.1016/j.ymeth.2016.04.025.

[111] Ying Wang, Craig L. Zirbel, Neocles B. Leontis, and Biao Ding. RNA 3-dimensional structural motifs as a critical constraint of viroid RNA evolution. *PLOS Pathogens*, 14(2):e1006801, February 2018. doi: 10.1371/journal.ppat.1006801. URL https://doi.org/10.1371/journal.ppat.1006801.

[112] K Y Chang and I Tinoco. Characterization of a "kissing" hairpin complex derived from the human immunodeficiency virus genome. *Proceedings of the National Academy of Sciences*, 91(18):8705–8709, August 1994. doi: 10.1073/pnas.91.18.8705. URL https://doi.org/10.1073/pnas.91.18.8705.

[113] Pierce Radecki, Rahul Uppuluri, Kaustubh Deshpande, and Sharon Aviran. Accurate detection of RNA stem-loops in structurome data reveals widespread association with protein binding sites. *RNA Biology*, 18(sup1):521–536, October 2021. doi: 10.1080/15476286.2021.1971382. URL https://doi.org/10.1080/15476286.2021.1971382.

[114] Chen Bao, Mingyi Zhu, Inna Nykonchuk, Hironao Wakabayashi, David H. Mathews, and Dmitri N. Ermolenko. Specific length and structure rather than high thermodynamic stability enable regulatory mRNA stem-loops to pause translation. *Nature Communications*, 13(1), February 2022. doi: 10.1038/s41467-022-28600-5. URL https://doi.org/10.1038/s41467-022-28600-5.

[115] Jin-Der Wen, Laura Lancaster, Courtney Hodges, Ana-Carolina Zeri, Shige H. Yoshimura, Harry F. Noller, Carlos Bustamante, and Ignacio Tinoco. Following translation by single ribosomes one codon at a time. *Nature*, 452(7187):598–603, March 2008. doi: 10.1038/nature06716. URL https://doi.org/10.1038/nature06716.

[116] Thomas A Geissmann and Danièle Touati. Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *The EMBO Journal*, 23(2):396–405, January 2004. doi: 10.1038/sj.emboj.7600058. URL https://doi.org/10.1038/sj.emboj.7600058.

[117] K. Rietveld, R. Van Poelgeest, C.W.A. Pleij, J.H. Van Boom, and L. Bosch. The tRNA-uke structure at the 3′ terminus of turnip yellow mosaic virus RNA. differences and similarities with canonical tRNA. *Nucleic Acids Research*, 10(6):1929–1946, 1982. doi: 10.1093/nar/10.6.1929. URL https://doi.org/10.1093/nar/10.6.1929.

[118] David W Staple and Samuel E Butcher. Pseudoknots: RNA structures with diverse functions. *PLoS Biology*, 3(6):e213, June 2005. doi: 10.1371/journal.pbio.0030213. URL https://doi.org/10.1371/journal.pbio.0030213.

[119] Martin Egli, George Minasov, Li Su, and Alexander Rich. Metal ions and flexibility in a viral RNA pseudoknot at atomic resolution. *Proceedings of the National Academy of Sciences*, 99(7):4302–4307, March 2002. doi: 10.1073/pnas.062055599. URL https://doi.org/10.1073/pnas.062055599.

[120] F. H. D. van Batenburg. PseudoBase: structural information on RNA pseudoknots. *Nucleic Acids Research*, 29(1):194–195, January 2001. doi: 10.1093/nar/29.1.194. URL https://doi.org/10.1093/nar/29.1.194.

[121] Ailong Ke, Kaihong Zhou, Fang Ding, Jamie H. D. Cate, and Jennifer A. Doudna. A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature*, 429(6988):201–205, May 2004. doi: 10.1038/nature02522. URL https://doi.org/10.1038/nature02522.

[122] T. Powers and H.F. Noller. A functional pseudoknot in 16s ribosomal RNA. *The EMBO Journal*, 10(8):2203–2214, August 1991. doi: 10.1002/j.1460-2075.1991.tb07756.x. URL https://doi.org/10.1002/j.1460-2075.1991.tb07756.x.

[123] Peter L. Adams, Mary R. Stahley, Anne B. Kosek, Jimin Wang, and Scott A. Strobel. Crystal structure of a self-splicing group i intron with both exons. *Nature*, 430(6995):45–50, June 2004. doi: 10.1038/nature02642. URL https://doi.org/10.1038/nature02642.

[124] Carla A. Theimer, Craig A. Blois, and Juli Feigon. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Molecular Cell*, 17(5):671–682, March 2005. doi: 10.1016/j.molcel.2005.01.017. URL https://doi.org/10.1016/j.molcel.2005.01.017.

[125] Jiunn-Liang Chen and Carol W. Greider. Functional analysis of the pseudoknot structure in human telomerase RNA. *Proceedings of the National Academy of Sciences*, 102(23):8080–8085, April 2005. doi: 10.1073/pnas.0502259102. URL https://doi.org/10.1073/pnas.0502259102.

[126] Ling X. Shen and Jr Ignacio Tinoco. The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *Journal of Molecular Biology*, 247(5):963–978, April 1995. doi: 10.1006/jmbi.1995.0193. URL https://doi.org/10.1006/jmbi.1995.0193.

[127] Paul L Nixon, Anupama Rangan, Y.-G Kim, Alexander Rich, David W Hoffman, Mirko Hennig, and David P Giedroc. Solution structure of a luteoviral p1–p2 frameshifting mRNA pseudoknot. *Journal of Molecular Biology*, 322(3):621–633, September 2002. doi: 10.1016/s0022-2836(02)00779-9. URL https://doi.org/10.1016/s0022-2836(02)00779-9.

[128] Paul J.A Michiels, Alexandra A.M Versleijen, Paul W Verlaan, Cornelis W.A Pleij, Cornelis W Hilbers, and Hans A Heus. Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting 1 1edited by i. tinoco. *Journal of Molecular Biology*, 310(5):1109–1123, July 2001. doi: 10.1006/jmbi.2001.4823. URL https://doi.org/10.1006/jmbi.2001.4823.

[129] Krishna Neupane, Meng Zhao, Aaron Lyons, Sneha Munshi, Sandaru M. Ileperuma, Dustin B. Ritchie, Noel Q. Hoffer, Abhishek Narayan, and Michael T. Woodside. Structural dynamics of single SARS-CoV-2 pseudoknot molecules reveal topologically distinct conformers. *Nature Communications*, 12(1), August 2021. doi: 10.1038/s41467-021-25085-6. URL https://doi.org/10.1038/s41467-021-25085-6.

[130] Christian Laing and Tamar Schlick. Analysis of four-way junctions in RNA structures. *Journal of Molecular Biology*, 390(3):547–559, July 2009. doi: 10.1016/j.jmb.2009.04.084. URL https://doi.org/10.1016/j.jmb.2009.04.084.

[131] Aurélie Lescoute and Eric Westhof. Topology of three-way junctions in folded RNAs. *RNA*, 12(1):83–93, December 2005. doi: 10.1261/rna.2208106. URL https://doi.org/10.1261/rna.2208106.

[132] Sunny C.Y. Jeng, Robert J. Trachman, Florian Weissenboeck, Lynda Truong, Katie A. Link, Mette D.E. Jepsen, Jay R. Knutson, Ebbe S. Andersen, Adrian R. Ferré-D'Amaré, and Peter J. Unrau. Fluorogenic aptamers resolve the flexibility of RNA junctions using orientation-dependent FRET. *RNA*, 27(4):433–444, December 2020. doi: 10.1261/rna.078220.120. URL https://doi.org/10.1261/rna.078220.120.

[133] Sarah Knight Denny, Namita Bisaria, Joseph David Yesselman, Rhiju Das, Daniel Herschlag, and William James Greenleaf. High-throughput investigation of diverse junction elements in RNA tertiary folding. *Cell*, 174(2):377–390.e20, July 2018. doi: 10.1016/j.cell.2018.05.038. URL https://doi.org/10.1016/j.cell.2018.05.038.

[134] S. H. Kim, J. L. Sussman, F. L. Suddath, G. J. Quigley, A. McPherson, A. H. J. Wang, N. C. Seeman, and Alexander Rich. The general structure of transfer RNA molecules. *Proceedings of the National Academy of Sciences*, 71(12):4970–4974, December 1974. doi: 10.1073/pnas.71.12.4970. URL https://doi.org/10.1073/pnas.71.12.4970.

[135] Stephen R Holbrook. RNA structure: the long and the short of it. *Current Opinion in Structural Biology*, 15(3):302–308, June 2005. doi: 10.1016/j.sbi.2005.04.005. URL https://doi.org/10.1016/j.sbi.2005.04.005.

[136] Stephen R. Holbrook. Structural principles from large RNAs. *Annual Review of Biophysics*, 37(1): 445–464, June 2008. doi: 10.1146/annurev.biophys.36.040306.132755. URL https://doi.org/10.1146/annurev.biophys.36.040306.132755.

[137] Yurong Xin, Christian Laing, Neocles B. Leontis, and Tamar Schlick. Annotation of tertiary interactions in RNA structures reveals variations and correlations. *RNA*, 14(12):2465–2477, October 2008. doi: 10.1261/rna.1249208. URL https://doi.org/10.1261/rna.1249208.

[138] Sungchul Hohng, Timothy J Wilson, Elliot Tan, Robert M Clegg, David M.J Lilley, and Taekjip Ha. Conformational flexibility of four-way junctions in RNA. *Journal of Molecular Biology*, 336(1):69–79, February 2004. doi: 10.1016/j.jmb.2003.12.014. URL https://doi.org/10.1016/j.jmb.2003.12.014.

[139] Anastasia Khvorova, Aurélie Lescoute, Eric Westhof, and Sumedha D Jayasena. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nature Structural & Molecular Biology*, 10(9):708–712, July 2003. doi: 10.1038/nsb959. URL https://doi.org/10.1038/nsb959.

[140] Marella D. Canny, Fiona M. Jucker, Elizabeth Kellogg, Anastasia Khvorova, Sumedha D. Jayasena, and Arthur Pardi. Fast cleavage kinetics of a natural hammerhead ribozyme. *Journal of the American Chemical Society*, 126(35):10848–10849, August 2004. doi: 10.1021/ja046848v. URL https://doi.org/10.1021/ja046848v.

[141] J. Carlos Penedo, Timothy J. Wilson, Sumedha D. Jayasena, Anastasia Khrova, and David M.J. Lilley. Folding of the natural hammerhead ribozyme is enhanced by interaction of auxiliary elements. *RNA*, 10(5):880–888, April 2004. doi: 10.1261/rna.5268404. URL https://doi.org/10.1261/rna.5268404.

[142] Francesco Di Palma, Francesco Colizzi, and Giovanni Bussi. Ligand-induced stabilization of the aptamer terminal helix in the adenine riboswitch. *RNA*, 19(11):1517–1524, September 2013. doi: 10.1261/rna.040493.113. URL https://doi.org/10.1261/rna.040493.113.

[143] Elliot Tan, Timothy J. Wilson, Michelle K. Nahas, Robert M. Clegg, David M. J. Lilley, and Taekjip Ha. A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. *Proceedings of the National Academy of Sciences*, 100(16):9308–9313, July 2003. doi: 10.1073/pnas.1233536100. URL https://doi.org/10.1073/pnas.1233536100.

[144] Wade C Winkler, Ali Nahvi, Narasimhan Sudarsan, Jeffrey E Barrick, and Ronald R Breaker. An mRNA structure that controls gene expression by binding s-adenosylmethionine. *Nature Structural & Molecular Biology*, 10(9):701–707, August 2003. doi: 10.1038/nsb967. URL https://doi.org/10.1038/nsb967.

[145] Makio Tamura and Stephen R Holbrook. Sequence and structural conservation in RNA ribose zippers. *Journal of Molecular Biology*, 320(3):455–474, July 2002. doi: 10.1016/s0022-2836(02)00515-6. URL https://doi.org/10.1016/s0022-2836(02)00515-6.

[146] Samuel E. Butcher and Anna Marie Pyle. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Accounts of Chemical Research*, 44(12):1302–1311, September 2011. doi: 10.1021/ar200098t. URL https://doi.org/10.1021/ar200098t.

[147] Christine Brunel, Roland Marquet, Pascale Romby, and Chantal Ehresmann. RNA loop–loop interactions as dynamic functional motifs. *Biochimie*, 84(9):925–944, September 2002. doi: 10.1016/s0300-9084(02)01401-3. URL https://doi.org/10.1016/s0300-9084(02)01401-3.

[148] Jennifer C. Darnell, Claire E. Fraser, Olga Mostovetsky, Giovanni Stefani, Thomas A. Jones, Sean R. Eddy, and Robert B. Darnell. Kissing complex RNAs mediate interaction between the fragile-x mental retardation protein KH2 domain and brain polyribosomes. *Genes & Development*, 19(8):903–918, April 2005. doi: 10.1101/gad.1276805. URL https://doi.org/10.1101/gad.1276805.

[149] Poul Nissen, Joseph A. Ippolito, Nenad Ban, Peter B. Moore, and Thomas A. Steitz. RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proceedings of the National Academy of Sciences*, 98(9):4899–4903, April 2001. doi: 10.1073/pnas.081082398. URL https://doi.org/10.1073/pnas.081082398.

[150] Donna K. Hendrix, Steven E. Brenner, and Stephen R. Holbrook. RNA structural motifs: building blocks of a modular biomolecule. *Quarterly Reviews of Biophysics*, 38(3):221–243, August 2005. doi: 10.1017/s0033583506004215. URL https://doi.org/10.1017/s0033583506004215.

[151] Luigi Bonati, Enrico Trizio, Andrea Rizzi, and Michele Parrinello. A unified framework for machine learning collective variables for enhanced sampling simulations: mlcolvar. *The Journal of Chemical Physics*, 159(1):014801, 2023.

[152] Anthony M. Mustoe, Charles L. Brooks, and Hashim M. Al-Hashimi. Hierarchy of RNA functional dynamics. *Annual Review of Biochemistry*, 83(1):441–466, 2014. doi: 10.1146/annurev-biochem-060713-035524. URL https://www.annualreviews.org/doi/10.1146/annurev-biochem-060713-035524. PMID: 24606137.

[153] Laura R. Ganser, Megan L. Kelly, Daniel Herschlag, and Hashim M. Al-Hashimi. The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology*, 20(8):474–489, June 2019. doi: 10.1038/s41580-019-0136-0. URL https://doi.org/10.1038/s41580-019-0136-0.

[154] Hashim M Al-Hashimi and Nils G Walter. RNA dynamics: it is about time. *Current Opinion in Structural Biology*, 18(3):321–329, June 2008. doi: 10.1016/j.sbi.2008.04.004. URL https://doi.org/10.1016/j.sbi.2008.04.004.

[155] Daniel Herschlag. RNA chaperones and the RNA folding problem. *Journal of Biological Chemistry*, 270(36):20871–20874, September 1995. doi: 10.1074/jbc.270.36.20871. URL https://doi.org/10.1074/jbc.270.36.20871.

[156] Manuela J. Rist and John P. Marino. Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry*, 41(50):14762–14770, November 2002. doi: 10.1021/bi0267240. URL https://doi.org/10.1021/bi0267240.

[157] David E. Shaw, Peter J. Adams, Asaph Azaria, Joseph A. Bank, Brannon Batson, Alistair Bell, Michael Bergdorf, Jhanvi Bhatt, J. Adam Butts, Timothy Correia, Robert M. Dirks, Ron O. Dror, Michael P. Eastwood, Bruce Edwards, Amos Even, Peter Feldmann, Michael Fenn, Christopher H. Fenton, Anthony Forte, Joseph Gagliardo, Gennette Gill, Maria Gorlatova, Brian Greskamp, J.P.

Grossman, Justin Gullingsrud, Anissa Harper, William Hasenplaugh, Mark Heily, Benjamin Colin Heshmat, Jeremy Hunt, Douglas J. Ierardi, Lev Iserovich, Bryan L. Jackson, Nick P. Johnson, Mollie M. Kirk, John L. Klepeis, Jeffrey S. Kuskin, Kenneth M. Mackenzie, Roy J. Mader, Richard McGowen, Adam McLaughlin, Mark A. Moraes, Mohamed H. Nasr, Lawrence J. Nociolo, Lief O'Donnell, Andrew Parker, Jon L. Peticolas, Goran Pocina, Cristian Predescu, Terry Quan, John K. Salmon, Carl Schwink, Keun Sup Shim, Naseer Siddique, Jochen Spengler, Tamas Szalay, Raymond Tabladillo, Reinhard Tartler, Andrew G. Taube, Michael Theobald, Brian Towles, William Vick, Stanley C. Wang, Michael Wazlowski, Madeleine J. Weingarten, John M. Williams, and Kevin A. Yuh. Anton 3. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. ACM, November 2021. doi: 10.1145/3458817.3487397. URL https://doi.org/10.1145/3458817.3487397.

[158] Pablo Herrera-Nieto, Adrià Pérez, and Gianni De Fabritiis. Characterization of partially ordered states in the intrinsically disordered N-terminal domain of p53 using millisecond Molecular Dynamics simulations. *Scientific Reports*, 10(1), July 2020. doi: 10.1038/s41598-020-69322-2. URL https://doi.org/10.1038/s41598-020-69322-2.

[159] Dazhi Tan, William F. Marzluff, Zbigniew Dominski, and Liang Tong. Structure of histone mRNA stem-loop, human stem-loop binding protein, and 3'hexo ternary complex. *Science*, 339(6117):318–321, 2013. doi: 10.1126/science.1228705. URL https://www.science.org/doi/abs/10.1126/science.1228705.

[160] Maximillian H. Bailor, Xiaoyan Sun, and Hashim M. Al-Hashimi. Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science*, 327(5962):202–206, January 2010. doi: 10.1126/science.1181085. URL https://doi.org/10.1126/science.1181085.

[161] Qi Zhang, Xiaoyan Sun, Eric D Watt, and Hashim M Al-Hashimi. Resolving the motional modes that code for RNA adaptation. *Science*, 311(5761):653–656, 2006. doi: 10.1126/science.1119488. URL https://www.science.org/doi/10.1126/science.1119488.

[162] Qi Zhang, Andrew C Stelzer, Charles K Fisher, and Hashim M Al-Hashimi. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature*, 450(7173):1263–1267, 2007. doi: https://doi.org/10.1038/nature06389. URL https://www.nature.com/articles/nature06389.

[163] Loïc Salmon, Gavin Bascom, Ioan Andricioaei, and Hashim M Al-Hashimi. A general method for constructing atomic-resolution RNA ensembles using NMR residual dipolar couplings: the basis for interhelical motions revealed. *Journal of the American Chemical Society*, 135(14):5457–5466, 2013. doi: https://doi.org/10.1021/ja400920w. URL https://pubs.acs.org/doi/full/10.1021/ja400920w.

[164] Grabow Wade and Jaeger Luc. RNA modularity for synthetic biology. *F1000Prime Rep*, 45(5), 2013. doi: https://doi.org/10.12703/P5-46. URL https://facultyopinions.com/prime/reports/b/5/46/.

[165] Chase L. Beisel, Yvonne Y. Chen, Stephanie J. Culler, Kevin G. Hoff, and Christina D. Smolke. Design of small molecule-responsive microRNAs based on structural requirements for Drosha processing. *Nucleic Acids Research*, 39(7):2981–2994, 12 2010. ISSN 0305-1048. doi: 10.1093/nar/gkq954. URL https://doi.org/10.1093/nar/gkq954.

[166] Jane S. Richardson, Bohdan Schneider, Laura W. Murray, Gary J. Kapral, Robert M. Immormino, Jeffrey J. Headd, David C. Richardson, Daniela Ham, Eli Hershkovits, Loren Dean Williams, Kevin S. Keating, Anna Marie Pyle, David Micallef, John Westbrook, and Helen M. Berman. RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA ontology consortium contribution). *RNA*, 14(3):465–481, January 2008. doi: 10.1261/rna.657708. URL https://doi.org/10.1261/rna.657708.

[167] Stephen C Harvey, M Prabhakaran, Boryeu Mao, and J Andrew McCammon. Phenylalanine transfer RNA: Molecular Dynamics simulation. *Science*, 223(4641):1189–1191, 1984. doi: 10.1126/science.6560785. URL https://www.science.org/doi/10.1126/science.6560785.

[168] Daniel R Roe, Christina Bergonzo, and Thomas E Cheatham III. Evaluation of enhanced sampling provided by accelerated Molecular Dynamics with Hamiltonian replica exchange methods. *The Journal of Physical Chemistry B*, 118(13):3543–3552, 2014. doi: https://doi.org/10.1021/jp4125099. URL https://pubs.acs.org/doi/10.1021/jp4125099.

[169] Christina Bergonzo, Niel M Henriksen, Daniel R Roe, and Thomas E Cheatham. Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA*, 21(9):1578–1590, 2015. doi: 10.1261/rna.051102.115. URL https://rnajournal.cshlp.org/content/21/9/1578.

[170] Pavel Banás, Daniel Hollas, Marie Zgarbová, Petra Jurecka, Modesto Orozco, Thomas E Cheatham III, Jiří Šponer, and Michal Otyepka. Performance of molecular mechanics force fields for RNA simulations: stability of UUCG and GNRA hairpins. *Journal of Chemical Theory and Computation*, 6(12):3836–3849, 2010. doi: 10.1021/ct100481h. URL https://pubs.acs.org/doi/10.1021/ct100481h.

[171] Ilyas Yildirim, Harry A Stern, Jason D Tubbs, Scott D Kennedy, and Douglas H Turner. Benchmarking AMBER force fields for RNA: Comparisons to NMR spectra for single-stranded rGACC) are improved by revised χ torsions. *The journal of physical chemistry B*, 115(29):9261–9270, 2011. doi: https://doi.org/10.1021/jp2016006. URL https://pubs.acs.org/doi/10.1021/jp2016006.

[172] David E Condon, Scott D Kennedy, Brendan C Mort, Ryszard Kierzek, Ilyas Yildirim, and Douglas H Turner. Stacking in RNA: NMR of four tetramers benchmark Molecular Dynamics. *Journal of chemical theory and computation*, 11(6):2729–2742, 2015. doi: 10.1021/ct501025q. URL https://pubs.acs.org/doi/10.1021/ct501025q.

[173] Petra Kuhrova, Robert B Best, Sandro Bottaro, Giovanni Bussi, Jiří Šponer, Michal Otyepka, and Pavel Banas. Computer folding of RNA tetraloops: identification of key force field deficiencies. *Journal of chemical theory and computation*, 12(9):4534–4548, 2016. doi: https://doi.org/10.1021/acs.jctc.6b00300. URL https://pubs.acs.org/doi/10.1021/acs.jctc.6b00300.

[174] Alessandra Villa and Gerhard Stock. What NMR relaxation can tell us about the internal motion of an RNA hairpin: a Molecular Dynamics simulation study. *Journal of chemical theory and computation*, 2(5):1228–1236, 2006. doi: 10.1021/ct600160z. URL https://pubs.acs.org/doi/10.1021/ct600160z.

[175] Eric J Sorin, Young Min Rhee, and Vijay S Pande. Does water play a structural role in the folding of small nucleic acids? *Biophysical journal*, 88(4):2516–2524, 2005. doi: https://doi.org/

10.1529/biophysj.104.055087. URL https://www.sciencedirect.com/science/article/pii/S0006349505733063.

[176] Eric J Sorin, Young Min Rhee, Bradley J Nakatani, and Vijay S Pande. Insights into nucleic acid conformational dynamics from massively parallel stochastic simulations. *Biophysical journal*, 85(2): 790–803, 2003. doi: 10.1016/S0006-3495(03)74520-2. URL https://www.sciencedirect.com/science/article/pii/S0006349503745202?via%3Dihub.

[177] Wenbing Zhang and Shi-Jie Chen. RNA hairpin-folding kinetics. *Proceedings of the National Academy of Sciences*, 99(4):1931–1936, 2002. doi: https://doi.org/10.1073/pnas.032443099. URL https://www.pnas.org/doi/10.1073/pnas.032443099.

[178] Xiaojun Xu, Tao Yu, and Shi-Jie Chen. Understanding the kinetic mechanism of RNA single base pair formation. *Proceedings of the National Academy of Sciences*, 113(1):116–121, 2016. doi: 10.1073/pnas.1705642114. URL https://www.pnas.org/doi/full/10.1073/pnas.1705642114.

[179] Angel E Garcia and Dietmar Paschek. Simulation of the pressure and temperature folding/unfolding equilibrium of a small RNA hairpin. *Journal of the American Chemical Society*, 130(3):815–817, 2008. doi: 10.1021/ja074191i. URL https://pubs.acs.org/doi/10.1021/ja074191i.

[180] Alessandra Villa, Elisabeth Widjajakusuma, and Gerhard Stock. Molecular Dynamics simulation of the structure, dynamics, and thermostability of the RNA hairpins uCACGg and cUUCGg. *The Journal of Physical Chemistry B*, 112(1):134–142, 2008. doi: 10.1021/jp0764337. URL https://pubs.acs.org/doi/10.1021/jp0764337.

[181] Alan A Chen and Angel E García. High-resolution reversible folding of hyperstable RNA tetraloops using Molecular Dynamics simulations. *Proceedings of the National Academy of Sciences*, 110(42): 16820–16825, 2013. doi: https://doi.org/10.1073/pnas.1309392110. URL https://www.pnas.org/doi/10.1073/pnas.1309392110.

[182] Peter A Kollman, Irina Massova, Carolina Reyes, Bernd Kuhn, Shuanghong Huo, Lillian Chong, Matthew Lee, Taisung Lee, Yong Duan, Wei Wang, et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of chemical research*, 33(12):889–897, 2000. doi: https://doi.org/10.1021/ar000033j. URL https://pubs.acs.org/doi/10.1021/ar000033j.

[183] Marek Havrila, Kamila Réblová, Craig L Zirbel, Neocles B Leontis, and Jiří Šponer. Isosteric and nonisosteric base pairs in RNA motifs: Molecular Dynamics and bioinformatics study of the sarcin–ricin internal loop. *The journal of physical chemistry B*, 117(46):14302–14319, 2013. doi: https://doi.org/10.1021/jp408530w. URL https://pubs.acs.org/doi/10.1021/jp408530w.

[184] Kamila Réblová, Nadá Špačková, Richard Štefl, Kristina Csaszar, Jaroslav Koča, Neocles B Leontis, and Jiří Šponer. Non-watson-crick basepairing and hydration in RNA motifs: Molecular Dynamics of 5s rRNA loop E. *Biophysical journal*, 84(6):3564–3582, 2003. doi: 10.1016/S0006-3495(03)75089-9. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1302943/.

[185] Pascal Auffinger, Lukasz Bielecki, and Eric Westhof. Symmetric K+ and Mg2+ ion-binding sites in the 5S rRNA loop E inferred from Molecular Dynamics simulations. *Journal of molecular biology*, 335(2): 555–571, 2004. doi: 10.1016/S0006-3495(03)75089-9. URL https://www.cell.com/biophysj/fulltext/S0006-3495(03)75089-9.

[186] Filip Rázga, Jaroslav Koča, Jiří Šponer, and Neocles B Leontis. Hinge-like motions in RNA kink-turns: the role of the second a-minor motif and nominally unpaired bases. *Biophysical journal*, 88(5): 3466–3485, 2005. doi: 10.1529/biophysj.104.054916. URL https://www.cell.com/biophysj/fulltext/S0006-3495(05)73395-6.

[187] Kamila Réblová, Nadá Špačková, Judit E Šponer, Jaroslav Koča, and Jiří Šponer. Molecular Dynamics simulations of RNA kissing–loop motifs reveal structural dynamics and formation of cation-binding pockets. *Nucleic acids research*, 31(23):6942–6952, 2003. doi: 10.1093/nar/gkg880. URL https://academic.oup.com/nar/article/31/23/6942/2904338.

[188] Tadeusz Kulinski, Mikolaj Olejniczak, Hendrik Huthoff, Lukasz Bielecki, Katarzyna Pachulska-Wieczorek, Atze T Das, Ben Berkhout, and Ryszard W Adamiak. The apical loop of the HIV-1 tar RNA hairpin is stabilized by a cross-loop base pair. *Journal of Biological Chemistry*, 278(40):38892–38901, 2003. doi: 10.1074/jbc.M301939200. URL https://www.jbc.org/article/S0021-9258(20)83202-5/fulltext.

[189] Elizabeth A Dethoff, Alexandar L Hansen, Catherine Musselman, Eric D Watt, Ioan Andricioaei, and Hashim M Al-Hashimi. Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, Molecular Dynamics, and mutagenesis. *Biophysical journal*, 95(8):3906–3915, 2008. URL https://linkinghub.elsevier.com/retrieve/pii/S0006349508785291.

[190] Brent A Gregersen, Xabier Lopez, and Darrin M York. Hybrid QM/MM study of thio effects in transphosphorylation reactions. *Journal of the American Chemical Society*, 125(24):7178–7179, 2003. doi: 10.1021/ja035167h. URL https://pubs.acs.org/doi/10.1021/ja035167h.

[191] Maryna V Krasovska, Jana Sefcikova, Kamila Réblová, Bohdan Schneider, Nils G Walter, and Jiří Šponer. Cations and hydration in catalytic RNA: Molecular Dynamics of the hepatitis delta virus ribozyme. *Biophysical journal*, 91(2):626–638, 2006. doi: 10.1529/biophysj.105.079368. URL https://www.sciencedirect.com/science/article/pii/S0006349506717611?via%3Dihub.

[192] Thomas Hermann, Pascal Auffinger, and Eric Westhof. Molecular Dynamics investigations of hammerhead ribozyme RNA. *European biophysics journal*, 27:153–165, 1998. doi: 10.1007/s002490050121. URL https://link.springer.com/article/10.1007/s002490050121.

[193] Stefan Trobro and Johan Åqvist. Analysis of predictions for the catalytic mechanism of ribosomal peptidyl transfer. *Biochemistry*, 45(23):7049–7056, 2006. doi: 10.1021/bi0605383. URL https://pubs.acs.org/doi/10.1021/bi0605383.

[194] Xiancheng Zeng, Jeetender Chugh, Anette Casiano-Negroni, Hashim M Al-Hashimi, and Charles L Brooks III. Flipping of the ribosomal A-site adenines provides a basis for tRNA selection. *Journal of molecular biology*, 426(19):3201–3213, 2014. doi: 10.1016/j.jmb.2014.04.029. URL https://www.sciencedirect.com/science/article/pii/S0022283614002307?via%3Dihub.

[195] Priyadarshi Satpati and Johan Åqvist. Why base tautomerization does not cause errors in mRNA decoding on the ribosome. *Nucleic acids research*, 42(20):12876–12884, 2014. doi: 10.1093/nar/gku1044. URL https://academic.oup.com/nar/article/42/20/12876/2902660?login=false.

[196] Stefan Trobro and Johan Åqvist. Mechanism of peptide bond synthesis on the ribosome. *Proceedings of the National Academy of Sciences*, 102(35):12395–12400, 2005. doi: 10.1073/pnas.0504043102. URL https://www.pnas.org/doi/10.1073/pnas.0504043102.

[197] Andrea C Vaiana and Kevin Y Sanbonmatsu. Stochastic gating and drug–ribosome interactions. *Journal of molecular biology*, 386(3):648–661, 2009. doi: 10.1016/j.jmb.2008.12.035. URL https://www.sciencedirect.com/science/article/pii/S0022283608015465?via%3Dihub.

[198] Julia Romanowska, Nathalie Reuter, and Joanna Trylska. Comparing aminoglycoside binding sites in bacterial ribosomal RNA and aminoglycoside modifying enzymes. *Proteins: Structure, Function, and Bioinformatics*, 81(1):63–80, 2013. doi: https://doi.org/10.1002/prot.24163. URL https://onlinelibrary.wiley.com/doi/full/10.1002/prot.24163.

[199] AC Vaiana, E Westhof, and P Auffinger. A Molecular Dynamics simulation study of an aminoglycoside/a-site RNA complex: conformational and hydration patterns. *Biochimie*, 88(8): 1061–1073, 2006. doi: 10.1016/j.biochi.2006.06.006. URL https://www.sciencedirect.com/science/article/pii/S0300908406001052?via%3Dihub.

[200] Kevin Y Sanbonmatsu, Simpson Joseph, and Chang-Shung Tung. Simulating movement of tRNA into the ribosome during decoding. *Proceedings of the National Academy of Sciences*, 102(44):15854–15859, 2005. doi: 10.1073/pnas.0503456102. URL https://www.pnas.org/doi/full/10.1073/pnas.0503456102.

[201] Lars V Bock, Christian Blau, Gunnar F Schröder, Iakov I Davydov, Niels Fischer, Holger Stark, Marina V Rodnina, Andrea C Vaiana, and Helmut Grubmüller. Energy barriers and driving forces in tRNA translocation through the ribosome. *Nature structural & molecular biology*, 20(12):1390–1396, 2013. doi: 10.1038/nsmb.2690. URL https://www.nature.com/articles/nsmb.2690.

[202] Paul C Whitford, Peter Geggier, Roger B Altman, Scott C Blanchard, José N Onuchic, and Karissa Y Sanbonmatsu. Accommodation of aminoacyl-tRNA into the ribosome involves reversible excursions along multiple pathways. *Rna*, 16(6):1196–1204, 2010. doi: 10.1261/rna.2035410. URL https://rnajournal.cshlp.org/content/16/6/1196.

[203] Leonardo G Trabuco, Eduard Schreiner, John Eargle, Peter Cornish, Taekjip Ha, Zaida Luthey-Schulten, and Klaus Schulten. The role of L1 stalk–tRNA interaction in the ribosome elongation cycle. *Journal of molecular biology*, 402(4):741–760, 2010. doi: 10.1016/j.jmb.2010.07.056. URL https://www.sciencedirect.com/science/article/pii/S0022283610008326?via%3Dihub.

[204] Chee-Hock Hoe, Carsten A. Raabe, Timofey S. Rozhdestvensky, and Thean-Hock Tang. Bacterial sRNAs: Regulation in stress. *International Journal of Medical Microbiology*, 303(5):217–229, July 2013. doi: 10.1016/j.ijmm.2013.04.002. URL https://doi.org/10.1016/j.ijmm.2013.04.002.

[205] Thorleif Møller, Thomas Franch, Peter Højrup, Douglas R Keene, Hans Peter Bächinger, Richard G Brennan, and Poul Valentin-Hansen. Hfq. *Molecular Cell*, 9(1):23–30, January 2002. doi: 10.1016/s1097-2765(01)00436-1. URL https://doi.org/10.1016/s1097-2765(01)00436-1.

[206] Mengxin Li, Yalong Cong, Yifei Qi, and John Z. H. Zhang. Computational insights into the binding mechanism of OxyS sRNA with chaperone protein hfq. *Biomolecules*, 11(11):1653, November 2021. doi: 10.3390/biom11111653. URL https://doi.org/10.3390/biom11111653.

[207] Xue Yuan Pei, Tom Dendooven, Elisabeth Sonnleitner, Shaoxia Chen, Udo Bläsi, and Ben F Luisi. Architectural principles for Hfq/Crc-mediated regulation of gene expression. *eLife*, 8, February 2019. doi: 10.7554/elife.43158. URL https://doi.org/10.7554/elife.43158.

[208] Prettina Lazar, Yuno Lee, Songmi Kim, Meganathan Chandrasekaran, and Keun Woo Lee. Molecular Dynamics simulation study for ionic strength dependence of RNA-host factor interaction in *staphylococcus aureus* HFQ. *Bulletin of The Korean Chemical Society*, 31:1519–1526, 2010. URL https://api.semanticscholar.org/CorpusID:4308697.

[209] Yong Zhang, Hisae Tateishi-Karimata, Tamaki Endoh, Qiongli Jin, Kexin Li, Xiaoru Fan, Yingjun Ma, Limin Gao, Haiyan Lu, Zhiye Wang, Art E. Cho, Xuefeng Yao, Chunming Liu, Naoki Sugimoto, Shiwei Guo, Xiangdong Fu, Qirong Shen, Guohua Xu, Luis Rafael Herrera-Estrella, and Xiaorong Fan. High-temperature adaptation of an OSNRT2.3 allele is thermoregulated by small RNAs. *Science Advances*, 8(47):eadc9785, 2022. doi: 10.1126/sciadv.adc9785. URL https://www.science.org/doi/abs/10.1126/sciadv.adc9785.

[210] Josiah Willard Gibbs. *Elementary principles in statistical mechanics: developed with especial reference to the rational foundations of thermodynamics*. C. Scribner's sons, 1902.

[211] James P. Sethna. 79Phase-space dynamics and ergodicity. In *Statistical Mechanics: Entropy, Order Parameters, and Complexity*. Oxford University Press, 01 2021. ISBN 9780198865247. doi: 10.1093/oso/9780198865247.003.0004. URL https://doi.org/10.1093/oso/9780198865247.003.0004.

[212] Ludwig Boltzmann. *Vorlesungen über Gastheorie: 2. Teil*. BoD–Books on Demand, 2017.

[213] Mark E. Tuckerman and Glenn J. Martyna. Understanding modern Molecular Dynamics: Techniques and applications. *The Journal of Physical Chemistry B*, 104(2):159–178, 2000. doi: 10.1021/jp992433y. URL https://doi.org/10.1021/jp992433y.

[214] Hiqmet Kamberaj. *Implementation of Molecular Dynamics in Simulations*, pages 167–182. Springer Nature Switzerland, 2023. ISBN 9783031348396. doi: 10.1007/978-3-031-34839-6_8. URL http://dx.doi.org/10.1007/978-3-031-34839-6_8.

[215] Dimitrios Vlachakis, Elena Bencurova, Nikitas Papangelopoulos, and Sophia Kossida. Current state-of-the-art Molecular Dynamics methods and applications. In *Advances in Protein Chemistry and Structural Biology*, pages 269–313. Elsevier, 2014. doi: 10.1016/b978-0-12-800168-4.00007-x. URL https://doi.org/10.1016/b978-0-12-800168-4.00007-x.

[216] Efrem Braun, Justin Gilmer, Heather B. Mayes, David L. Mobley, Jacob I. Monroe, Samarjeet Prasad, and Daniel M. Zuckerman. Best practices for foundations in molecular simulations [article v1.0]. *Living Journal of Computational Molecular Science*, 1(1), 2019. doi: 10.33011/livecoms.1.1.5957. URL https://doi.org/10.33011/livecoms.1.1.5957.

[217] Juliana Palma and Gustavo Pierdominici Sottile. Dinámica molecular de biomoléculas. *Divulgatio. Perfiles académicos de posgrado*, 7(20):63–99, May 2023. doi: 10.48160/25913530di20.292. URL https://doi.org/10.48160/25913530di20.292.

[218] Yaser Hashem and Pascal Auffinger. A short guide for Molecular Dynamics simulations of RNA systems. *Methods*, 47(3):187–197, 2009. ISSN 1046-2023. doi: https://doi.org/10.1016/j.ymeth.2008.09.020. URL https://www.sciencedirect.com/science/article/pii/S1046202308001576. Nucleic Acid Structure.

[219] John M. Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A A Kohl, Andy Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David A. Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with Alphafold. *Nature*, 596:583–589, 2021. doi: https://doi.org/10.1038/s41586-021-03819-2. URL https://www.nature.com/articles/s41586-021-03819-2.

[220] Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, Augustin Žídek, Tim Green, Kathryn Tunyasuvunakool, Stig Petersen, John Jumper, Ellen Clancy, Richard Green, Ankur Vora, Mira Lutfi, Michael Figurnov, Andrew Cowie, Nicole Hobbs, Pushmeet Kohli, Gerard Kleywegt, Ewan Birney, Demis Hassabis, and Sameer Velankar. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1):D439–D444, 11 2021. doi: https://doi.org/10.1093/nar/gkab1061. URL https://academic.oup.com/nar/article/50/D1/D439/6430488?login=false.

[221] B. Stahl. Alphafold server, 2023. URL https://www.alphafold.ebi.ac.uk/.

[222] H. M. Berman. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, January 2000. doi: 10.1093/nar/28.1.235. URL https://doi.org/10.1093/nar/28.1.235.

[223] Steven L Salzberg. Open questions: How many genes do we have? *BMC biology*, 16(1):1–3, 2018. doi: 10.1186/s12915-018-0564-x. URL https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-018-0564-x.

[224] Sergey Nurk, Sergey Koren, Arang Rhie, Mikko Rautiainen, Andrey V Bzikadze, Alla Mikheenko, Mitchell R Vollger, Nicolas Altemose, Lev Uralsky, Ariel Gershman, et al. The complete sequence of a human genome. *Science*, 376(6588):44–53, 2022. doi: 10.1126/science.abj6987. URL https://www.science.org/doi/10.1126/science.abj6987.

[225] Eric Paquet and Herna L. Viktor. Molecular Dynamics, Monte Carlo simulations, and Langevin Dynamics: A computational review. *BioMed Research International*, 2015:1–18, 2015. doi: 10.1155/2015/183918. URL https://doi.org/10.1155/2015/183918.

[226] David S. Cerutti and David A. Case. Molecular Dynamics simulations of macromolecular crystals. *WIREs Computational Molecular Science*, 9(4), November 2018. doi: 10.1002/wcms.1402. URL https://doi.org/10.1002/wcms.1402.

[227] Juan Carlos Jauregui. *Introduction*, pages 1–17. Elsevier, 2014. doi: 10.1533/9781782421665.1. URL http://dx.doi.org/10.1533/9781782421665.1.

[228] Ben Leimkuhler, Emad Noorizadeh, and Florian Theil. A gentle stochastic thermostat for Molecular Dynamics. *Journal of Statistical Physics*, 135(2):261–277, April 2009. doi: 10.1007/s10955-009-9734-0. URL https://doi.org/10.1007/s10955-009-9734-0.

[229] Herbert B. Callen and Theodore A. Welton. Irreversibility and generalized noise. *Physical Review*, 83 (1):34–40, July 1951. ISSN 0031-899X. doi: 10.1103/physrev.83.34. URL http://dx.doi.org/10.1103/PhysRev.83.34.

[230] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular Dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, October 1984. doi: 10.1063/1.448118. URL https://doi.org/10.1063/1.448118.

[231] Venkat Kapil, David M. Wilkins, Jinggang Lan, and Michele Ceriotti. Inexpensive modeling of quantum dynamics using path integral generalized langevin equation thermostats. *The Journal of Chemical Physics*, 152(12), March 2020. ISSN 1089-7690. doi: 10.1063/1.5141950. URL http://dx.doi.org/10.1063/1.5141950.

[232] P.W.M. Jacobs and Z.A. Rycerz. M*olecular* D*ynamics Methods*, pages 83–115. Elsevier, 1997. doi: 10.1016/b978-012164135-1/50005-8. URL http://dx.doi.org/10.1016/b978-012164135-1/50005-8.

[233] D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, III T.E. Cheatham, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York, and P.A. Kollman. Amber 18. *University of California, San Francisco*, 2018. URL https://ambermd.org/doc12/Amber18.pdf.

[234] R. Elber, A. P. Ruymgaart, and B. Hess. Shake parallelization. *The European Physical Journal Special Topics*, 200(1):211–223, November 2011. ISSN 1951-6401. doi: 10.1140/epjst/e2011-01525-9. URL http://dx.doi.org/10.1140/epjst/e2011-01525-9.

[235] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J.C Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular Dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, March 1977. ISSN 0021-9991. doi: 10.1016/0021-9991(77)90098-5. URL http://dx.doi.org/10.1016/0021-9991(77)90098-5.

[236] E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Physical Review*, 28(6):1049–1070, December 1926. ISSN 0031-899X. doi: 10.1103/physrev.28.1049. URL http://dx.doi.org/10.1103/PhysRev.28.1049.

[237] Louis G. Smith, Jianbo Zhao, David H. Mathews, and Douglas H. Turner. Physics-based all-atom modeling of RNA energetics and structure. *Wiley Interdisciplinary Reviews: RNA*, 8(5):e1422, August 2017. doi: 10.1002/wrna.1422. URL https://doi.org/10.1002/wrna.1422.

[238] E. Gerjuoy. Low-energy electron-atom and electron-molecule scattering theory circa 1964. *Physics Today*, 18(5):24–30, May 1965. ISSN 1945-0699. doi: 10.1063/1.3047421. URL http://dx.doi.org/10.1063/1.3047421.

[239] Lei Sun and Wei-Qiao Deng. Recent developments of first-principles force fields. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 7(1):e1282, September 2016. doi: 10.1002/wcms.1282. URL https://doi.org/10.1002/wcms.1282.

[240] Dazhi Tan, Stefano Piana, Robert M. Dirks, and David E. Shaw. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proceedings of the National Academy of Sciences*, 115(7), January 2018. doi: 10.1073/pnas.1713027115. URL https://doi.org/10.1073/pnas.1713027115.

[241] Debayan Chakraborty, Rosana Collepardo-Guevara, and David J Wales. Energy landscapes, folding mechanisms, and kinetics of RNA tetraloop hairpins. *Journal of the American Chemical Society*, 136 (52):18052–18061, 2014. doi: https://doi.org/10.1021/ja5100756. URL https://pubs.acs.org/doi/10.1021/ja5100756.

[242] Thomas E Cheatham and Peter A Kollman. Molecular Dynamics simulations highlight the structural differences among DNA: DNA, RNA: RNA, and DNA: RNA hybrid duplexes. *Journal of the American Chemical Society*, 119(21):4805–4825, 1997. doi: https://doi.org/10.1021/ja963641w. URL https://pubs.acs.org/doi/abs/10.1021/ja963641w.

[243] Sweta Vangaveti, Srivathsan V. Ranganathan, and Alan A. Chen. Advances in RNA Molecular Dynamics: a simulator's guide to RNA force fields. *WIREs RNA*, 8(2):e1396, 2017. doi: 10.1002/wrna.1396. URL https://wires.onlinelibrary.wiley.com/doi/10.1002/wrna.1396.

[244] H. A. Lorentz. Ueber die anwendung des satzes vom virial in der kinetischen theorie der gase. *Annalen der Physik*, 248(1):127–136, January 1881. ISSN 1521-3889. doi: 10.1002/andp.18812480110. URL http://dx.doi.org/10.1002/andp.18812480110.

[245] E. Lennard-Jones. On the determination of molecular fields. II. From the equation of state of a gas. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 106(738):463–477, October 1924. ISSN 2053-9150. doi: 10.1098/rspa.1924.0082. URL http://dx.doi.org/10.1098/rspa.1924.0082.

[246] P. P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921. doi: 10.1002/andp.19213690304. URL https://doi.org/10.1002/andp.19213690304.

[247] Thijs van Westen, Thijs J. H. Vlugt, and Joachim Gross. Determining force field parameters using a physically based equation of state. *The Journal of Physical Chemistry B*, 115(24):7872–7880, May 2011. ISSN 1520-5207. doi: 10.1021/jp2026219. URL http://dx.doi.org/10.1021/jp2026219.

[248] Michael S. Jones, Brennan Ashwood, Andrei Tokmakoff, and Andrew L. Ferguson. Determining sequence-dependent DNA oligonucleotide hybridization and dehybridization mechanisms using coarse-grained molecular simulation, Markov State Models, and infrared spectroscopy. *Journal of the American Chemical Society*, 143(42):17395–17411, October 2021. doi: 10.1021/jacs.1c05219. URL https://doi.org/10.1021/jacs.1c05219.

[249] Sereina Riniker. Fixed-charge atomistic force fields for Molecular Dynamics simulations in the condensed phase: An overview. *Journal of Chemical Information and Modeling*, 58(3):565–578, March 2018. doi: 10.1021/acs.jcim.8b00042. URL https://doi.org/10.1021/acs.jcim.8b00042.

[250] V.P. Gupta. Characterization of chemical reactions. In *Principles and Applications of Quantum Chemistry*, pages 385–433. Elsevier, 2016. doi: 10.1016/b978-0-12-803478-1.00012-1. URL https://doi.org/10.1016/b978-0-12-803478-1.00012-1.

[251] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An n.log(n) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993. doi: 10.1063/1.464397. URL https://doi.org/10.1063/1.464397.

[252] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth Particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995. doi: 10.1063/1.470117. URL https://doi.org/10.1063/1.470117.

[253] E. DubouÃl'-Dijon, M. Javanainen, P. Delcroix, P. Jungwirth, and H. Martinez-Seara. A practical guide to biologically relevant molecular simulations with charge scaling for electronic polarization. *The Journal of Chemical Physics*, 153(5), August 2020. ISSN 1089-7690. doi: 10.1063/5.0017775. URL http://dx.doi.org/10.1063/5.0017775.

[254] Alice E. A. Allen, Michael J. Robertson, Michael C. Payne, and Daniel J. Cole. Development and validation of the quantum mechanical bespoke protein force field. *ACS Omega*, 4(11):14537–14550, August 2019. ISSN 2470-1343. doi: 10.1021/acsomega.9b01769. URL http://dx.doi.org/10.1021/acsomega.9b01769.

[255] Jiří Šponer, Judit E. Šponer, Arnošt Mládek, Petr Jurečka, Pavel Banáš, and Michal Otyepka. Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics and experiment. *Biopolymers*, pages n/a–n/a, June 2013. doi: 10.1002/bip.22322. URL https://doi.org/10.1002/bip.22322.

[256] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179âĂŞ5197, May 1995. ISSN 1520-5126. doi: 10.1021/ja00124a002. URL http://dx.doi.org/10.1021/ja00124a002.

[257] Alberto Pérez, Iván Marchán, Daniel Svozil, Jiri Sponer, Thomas E Cheatham, 3rd, Charles A Laughton, and Modesto Orozco. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J*, 92(11):3817–3829, March 2007. URL https://www.sciencedirect.com/science/article/pii/S0006349507711827.

[258] Ivan Ivani, Pablo D Dans, Agnes Noy, Alberto Pérez, Ignacio Faustino, Adam Hospital, Jürgen Walther, Pau Andrio, Ramon Goñi, Alexandra Balaceanu, et al. Parmbsc1: a refined force field for DNA simulations. *Nature methods*, 13(1):55–58, 2016. doi: https://doi.org/10.1038/nmeth.3658. URL https://www.nature.com/articles/nmeth.3658.

[259] Ilyas Yildirim, Harry A Stern, Scott D Kennedy, Jason D Tubbs, and Douglas H Turner. Reparameterization of RNA χ torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine. *Journal of chemical theory and computation*, 6(5):1520–1531, 2010. doi: 10.1021/ct900604a. URL https://pubs.acs.org/doi/10.1021/ct900604a.

[260] Ilyas Yildirim, Scott D. Kennedy, Harry A. Stern, James M. Hart, Ryszard Kierzek, and Douglas H. Turner. Revision of AMBER torsional parameters for RNA improves free energy predictions for tetramer duplexes with GC and iGiC base pairs. *Journal of Chemical Theory and Computation*, 8(1):172–181, 2012. doi: 10.1021/ct200557r. URL https://pubs.acs.org/doi/10.1021/ct200557r. PMID: 22249447.

[261] Marie Zgarbová, Michal Otyepka, Jiří Šponer, Arnošt Mládek, Pavel Banáš, Thomas E. Cheatham, and Petr Jurečka. Refinement of the Cornell *et al.* nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *Journal of Chemical Theory and Computation*, 7(9):2886–2902, August 2011. doi: 10.1021/ct200162x. URL https://doi.org/10.1021/ct200162x.

[262] Weiwei He, Nawavi Naleem, Diego Kleiman, and Serdal Kirmizialtin. Refining the RNA force field with small-angle X-ray scattering of helix-junction-helix RNA. *The Journal of Physical Chemistry*

Letters, 13(15):3400–3408, April 2022. ISSN 1948-7185. doi: 10.1021/acs.jpclett.2c00359. URL http://dx.doi.org/10.1021/acs.jpclett.2c00359.

[263] Jianzhong Chen, Qingkai Zeng, Wei Wang, Haibo Sun, and Guodong Hu. Decoding the identification mechanism of an SAM-III riboswitch on ligands through multiple independent gaussian-accelerated Molecular Dynamics simulations. *Journal of Chemical Information and Modeling*, 62(23):6118–6132, November 2022. doi: 10.1021/acs.jcim.2c00961. URL https://doi.org/10.1021/acs.jcim.2c00961.

[264] Maxwell R. Tucker, Stefano Piana, Dazhi Tan, Michael V. LeVine, and David E. Shaw. Development of force field parameters for the simulation of single- and double-stranded DNA molecules and dna-protein complexes. *The Journal of Physical Chemistry B*, 126(24):4442–4457, June 2022. ISSN 1520-5207. doi: 10.1021/acs.jpcb.1c10971. URL http://dx.doi.org/10.1021/acs.jpcb.1c10971.

[265] Zhengxin Li, Junxi Mu, Jun Chen, and Hai-Feng Chen. Base-specific RNA force field improving the dynamics conformation of nucleotide. *International Journal of Biological Macromolecules*, 222: 680–690, December 2022. ISSN 0141-8130. doi: 10.1016/j.ijbiomac.2022.09.183. URL http://dx.doi.org/10.1016/j.ijbiomac.2022.09.183.

[266] Andrea Cesari, Sandro Bottaro, Kresten Lindorff-Larsen, Pavel Banáš, Jiřó šponer, and Giovanni Bussi. Fitting corrections to an RNA force field using experimental data. *Journal of Chemical Theory and Computation*, 15(6):3425–3431, May 2019. ISSN 1549-9626. doi: 10.1021/acs.jctc.9b00206. URL http://dx.doi.org/10.1021/acs.jctc.9b00206.

[267] Vojtěch Mlýnský, Petra Kührová, Tomáš Kühr, Michal Otyepka, Giovanni Bussi, Pavel Banáš, and Jiřó šponer. Fine-tuning of the amber RNA force field with a new term adjusting interactions of terminal nucleotides. *Journal of Chemical Theory and Computation*, 16(6):3936–3946, May 2020. ISSN 1549-9626. doi: 10.1021/acs.jctc.0c00228. URL http://dx.doi.org/10.1021/acs.jctc.0c00228.

[268] Michael J. Robertson, Yue Qian, Matthew C. Robinson, Julian Tirado-Rives, and William L. Jorgensen. Development and testing of the opls-aa/m force field for rna. *Journal of Chemical Theory and Computation*, 15(4):2734–2742, February 2019. ISSN 1549-9626. doi: 10.1021/acs.jctc.9b00054. URL http://dx.doi.org/10.1021/acs.jctc.9b00054.

[269] Christina Bergonzo and Thomas E. III Cheatham. Improved force field parameters lead to a better description of RNA structure. *Journal of Chemical Theory and Computation*, 11(9):3969–3972, 2015. doi: 10.1021/acs.jctc.5b00444. URL https://pubs.acs.org/doi/abs/10.1021/ja963641w. PMID: 26575892.

[270] Petra Kührová, Michal Otyepka, Jiří Šponer, and Pavel Banáš. Are waters around RNA more than just a solvent? an insight from Molecular Dynamics simulations. *Journal of Chemical Theory and Computation*, 10(1):401–411, 2014. doi: 10.1021/ct400663s. URL https://pubs.acs.org/doi/10.1021/ct400663s. PMID: 26579919.

[271] Daniel J Price and Charles L Brooks III. A modified TIP3P water potential for simulation with Ewald summation. *The Journal of chemical physics*, 121(20):10096–10103, 2004. doi: https://doi.org/10.1063/1.1808117. URL https://pubs.aip.org/aip/jcp/article/121/20/10096/532393/A-modified-TIP3P-water-potential-for-simulation.

[272] Saeed Izadi, Ramu Anandakrishnan, and Alexey V. Onufriev. Building water models: A different approach. *The Journal of Physical Chemistry Letters*, 5(21):3863âĂŞ3871, October 2014. ISSN 1948-7185. doi: 10.1021/jz501780a. URL http://dx.doi.org/10.1021/jz501780a.

[273] Yeyue Xiong, Parviz Seifpanahi Shabane, and Alexey V. Onufriev. Melting points of opc and opc3 water models. *ACS Omega*, 5(39):25087âĂŞ25094, September 2020. ISSN 2470-1343. doi: 10.1021/acsomega.0c02638. URL http://dx.doi.org/10.1021/acsomega.0c02638.

[274] Michael Feig. Computational protein structure refinement: almost there, yet still so far to go. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 7(3):e1307, March 2017. doi: 10.1002/wcms.1307. URL https://doi.org/10.1002/wcms.1307.

[275] Muhammad Taimoor Khan, , and Ayesha Ameen and. The lack of a centralized and universal tool for input file generation for Molecular Dynamics simulations and the development of such a tool: A review. *BioScientific Review*, 01(01):48–52, February 2019. doi: 10.32350/bsr.0101.06. URL https://doi.org/10.32350/bsr.0101.06.

[276] Hans C. Andersen. Molecular Dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, February 1980. doi: 10.1063/1.439486. URL https://doi.org/10.1063/1.439486.

[277] Qia Ke, Xiaoting Gong, Shouwei Liao, Chongxiong Duan, and Libo Li. Effects of thermostats/barostats on physical properties of liquids by Molecular Dynamics simulations. *Journal of Molecular Liquids*, 365:120116, November 2022. doi: 10.1016/j.molliq.2022.120116. URL https://doi.org/10.1016/j.molliq.2022.120116.

[278] Kenneth Lam, Koushik Kasavajhala, Sarah Gunasekera, and Carlos Simmerling. Accelerating the ensemble convergence of RNA hairpin simulations with a replica exchange structure reservoir. *Journal of Chemical Theory and Computation*, 18(6):3930–3947, 2022. doi: 10.1021/acs.jctc.2c00065. URL https://pubs.acs.org/doi/10.1021/acs.jctc.2c00065. PMID: 35502992.

[279] D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, and P.A. Kollman. *AMBER 16*. University of California, San Francisco, California, 2016. URL https://ambermd.org/doc12/Amber18.pdf.

[280] Rodrigo Cossio-Pérez, Gustavo Pierdominici-Sottile, Pablo Sobrado, and Juliana Palma. Molecular Dynamics simulations of substrate release from T*rypanosoma cruzi* UDP-galactopyranose mutase. *Journal of Chemical Information and Modeling*, 59(2):809–817, January 2019. doi: 10.1021/acs.jcim.8b00675. URL https://doi.org/10.1021/acs.jcim.8b00675.

[281] Gustavo Pierdominici-Sottile, Luciano Moffatt, and Juliana Palma. The dynamic behavior of the P2X4 ion channel in the closed conformation. *Biophysical Journal*, 111(12):2642–2650, December 2016. doi: 10.1016/j.bpj.2016.10.027. URL https://doi.org/10.1016/j.bpj.2016.10.027.

[282] Rafael C. Bernardi, Marcelo C.R. Melo, and Klaus Schulten. Enhanced sampling techniques in Molecular Dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1850(5):872–877, 2015. ISSN 0304-4165. doi: https://doi.org/10.1016/j.bbagen.2014.10.019.

URL https://www.sciencedirect.com/science/article/pii/S0304416514003559. Recent developments of Molecular Dynamics.

[283] Johannes Kästner. Umbrella Sampling. *WIREs Computational Molecular Science*, 1(6):932–942, 2011. doi: https://doi.org/10.1002/wcms.66. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.66.

[284] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. I. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992. doi: https://doi.org/10.1002/jcc.540130812. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540130812.

[285] Edina Rosta and Gerhard Hummer. Free energies from Dynamic Weighted Histogram Analysis using unbiased Markov State Model. *Journal of Chemical Theory and Computation*, 11(1):276–285, 2015. doi: 10.1021/ct500719p. URL https://doi.org/10.1021/ct500719p. PMID: 26574225.

[286] Ilsoo Kim and Toby W. Allen. Bennett's acceptance ratio and histogram analysis methods enhanced by Umbrella Sampling along a reaction coordinate in configurational space. *The Journal of Chemical Physics*, 136(16):164103, 2012. doi: 10.1063/1.3701766. URL https://doi.org/10.1063/1.3701766.

[287] Leandro Martínez. Automatic identification of mobile and rigid substructures in Molecular Dynamics simulations and fractional structural fluctuation analysis. *PLOS ONE*, 10(3):e0119264, March 2015. doi: 10.1371/journal.pone.0119264. URL https://doi.org/10.1371/journal.pone.0119264.

[288] Juliana Palma and Gustavo Pierdominici-Sottile. On the uses of PCA to characterise Molecular Dynamics simulations of biological macromolecules: Basics and tips for an effective use. *ChemPhysChem*, 24(2):e202200491, 2023. doi: https://doi.org/10.1002/cphc.202200491. URL https://chemistry-europe.onlinelibrary.wiley.com/doi/full/10.1002/cphc.202200491.

[289] Gustavo Pierdominici-Sottile and Juliana Palma. New insights into the meaning and usefulness of principal component analysis of concatenated trajectories. *Journal of computational chemistry*, 36 (7):424-432, March 2015. ISSN 0192-8651. doi: 10.1002/jcc.23811. URL https://doi.org/10.1002/jcc.23811.

[290] Angel E. García. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.*, 68:2696–2699, Apr 1992. doi: 10.1103/PhysRevLett.68.2696. URL https://link.aps.org/doi/10.1103/PhysRevLett.68.2696.

[291] Andrea Amadei, Antonius BM Linssen, and Herman JC Berendsen. Essential dynamics of proteins. *Proteins: Struct. Funct. Genet.*, 17(4):412–425, 1993.

[292] Isabella Daidone and Andrea Amadei. Essential dynamics: foundation and applications. *WIREs Computational Molecular Science*, 2(5):762–770, 2012. doi: https://doi.org/10.1002/wcms.1099. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1099.

[293] Florian Sittel and Gerhard Stock. Robust density-based clustering to identify metastable conformational states of proteins. *Journal of Chemical Theory and Computation*, 12(5):2426–2435, April 2016. doi: 10.1021/acs.jctc.5b01233. URL https://doi.org/10.1021/acs.jctc.5b01233.

[294] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14(1):33–38, February 1996. doi: 10.1016/0263-7855(96)00018-5. URL https://doi.org/10.1016/0263-7855(96)00018-5.

[295] Florian Sittel, Abhinav Jain, and Gerhard Stock. Principal component analysis of Molecular Dynamics: On the use of cartesian vs. internal coordinates. *The Journal of Chemical Physics*, 141(1), July 2014. ISSN 1089-7690. doi: 10.1063/1.4885338. URL http://dx.doi.org/10.1063/1.4885338.

[296] Laura Riccardi, Phuong H. Nguyen, and Gerhard Stock. Free-energy landscape of RNA hairpins constructed via dihedral angle principal component analysis. *The Journal of Physical Chemistry B*, 113 (52):16660–16668, 2009. doi: 10.1021/jp9076036. URL https://doi.org/10.1021/jp9076036. PMID: 20028141.

[297] Xin Jin and Jiawei Han. *K-Means Clustering*, pages 563–564. Springer US, 2011. ISBN 9780387301648. doi: 10.1007/978-0-387-30164-8_425. URL http://dx.doi.org/10.1007/978-0-387-30164-8_425.

[298] James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 2015. doi: 10.1021/acs.jctc.5b00255. URL https://doi.org/10.1021/acs.jctc.5b00255. PMID: 26574453.

[299] In Suk Joung and Thomas E. Cheatham. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *The Journal of Physical Chemistry B*, 112(30): 9020–9041, July 2008. doi: 10.1021/jp8001614. URL https://doi.org/10.1021/jp8001614.

[300] Andreas W. Götz, Mark J. Williamson, Dong Xu, Duncan Poole, Scott Le Grand, and Ross C. Walker. Routine microsecond molecular dynamics simulations with amber on gpus. 1. generalized born. *Journal of Chemical Theory and Computation*, 8(5):1542âĂŞ1555, April 2012. ISSN 1549-9626. doi: 10.1021/ct200909j. URL http://dx.doi.org/10.1021/ct200909j.

[301] Asaminew H. Aytenfisu, Aleksandar Spasic, Alan Grossfield, Harry A. Stern, and David H. Mathews. Revised RNA dihedral parameters for the AMBER force field improve RNA Molecular Dynamics. *Journal of Chemical Theory and Computation*, 13(2):900–915, 2017. doi: 10.1021/acs.jctc.6b00870. URL https://doi.org/10.1021/acs.jctc.6b00870. PMID: 28048939.

[302] Rodrigo Pérez-Cossio, Juliana Palma, and Gustavo Pierdominici-Sottile. Consistent principal component modes from Molecular Dynamics simulations of proteins. *Journal of Chemical Information and Modeling*, 57(4):826–834, 2017. doi: 10.1021/acs.jcim.6b00646. URL https://doi.org/10.1021/acs.jcim.6b00646. PMID: 28301154.

[303] Florian Sittel and Gerhard Stock. Perspective: Identification of collective variables and metastable states of protein dynamics. *The Journal of Chemical Physics*, 149(15):150901, October 2018. doi: 10.1063/1.5049637. URL https://doi.org/10.1063/1.5049637.

[304] Andrea Amadei, Antonius B. M. Linssen, and Herman J. C. Berendsen. Essential dynamics of proteins. *Proteins: Structure, Function, and Bioinformatics*, 17(4):412âĂŞ425, December 1993. ISSN 1097-0134. doi: 10.1002/prot.340170408. URL http://dx.doi.org/10.1002/prot.340170408.

[305] Oliver F. Lange, Nils-Alexander Lakomek, Christophe Farés, Gunnar F. Schröder, Korvin F. A. Walter, Stefan Becker, Jens Meiler, Helmut Grubmüller, Christian Griesinger, and Bert L. de Groot. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science*, 320(5882):1471–1475, June 2008. doi: 10.1126/science.1157092. URL https://doi.org/10.1126/science.1157092.

[306] Andrew Pohorille, Christopher Jarzynski, and Christophe Chipot. Good practices in free-energy calculations. *The Journal of Physical Chemistry B*, 114(32):10235–10253, 2010. doi: 10.1021/jp102971x. URL https://doi.org/10.1021/jp102971x. PMID: 20701361.

[307] Karine Lapouge, Elena Sineva, Magnus Lindell, Katja Starke, Carol S. Baker, Paul Babitzke, and Dieter Haas. Mechanism of *hcn*A mRNA recognition in the Gac/Rsm signal transduction pathway of P*seudomonas fluorescens*. *Molecular Microbiology*, 66(2):341–356, 2007. doi: https://doi.org/10.1111/j.1365-2958.2007.05909.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2007.05909.x.

[308] Robin Pearce, Gilbert S. Omenn, and Yang Zhang. De novo RNA tertiary structure prediction at atomic resolution using geometric potentials from deep learning. *(Preprint)*, May 2022. doi: 10.1101/2022.05.15.491755. URL https://doi.org/10.1101/2022.05.15.491755.

Tabla de siglas

Sigla	Significado
ARN	Ácido/s ribonucleico/s
ARNm	ARN mensajero
ARNr	ARN ribosomal
ARNp	ARN pequeño no codificante
ARNt	ARN de transferencia
CF	Campo/s de Fuerza
Cryo-EM	Criomicroscopía electrónica
DM	Dinámica Molecular
EE	Espacio Esencial
FDP	Función de Densidad de Probabilidad
IA	Inteligencia Artificial
Kd	Constante/s de disociación
MU	Motivo de Unión (Para RsmZ y *hcnA)*
PCR	Reacción en Cadena de la Polimerasa
PEL	Perfil de Energía Libre
RMN	Resonancia Magnética Nuclear
RMSD	Desviación cuadrática media
RMSF	Fluctuación cuadrática media
SC	Segmento simple cadena entre SL2 y SL3
SEL	Superficie/s de Energía Libre
SEP	Superficie de Energía Potencial
SL	*Stem-loop*
SU	Sitio de unión (para RsmE)