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

Universidad
Nacional
de Quilmes

Grille Coronel, Leandro

Estados conformacionales ultracompactos de proteínas nativas

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:
Grille Coronel, L. (2019). Estados conformacionales ultracompactos de proteínas nativas. (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/20.500.11807/2054

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

Leandro Grille Coronel, Repositorio Institucional Digital de Acceso Abierto,
Mayo de 2019, pp. 115,
http://ridaa.unq.edu.ar,
Universidad Nacional de Quilmes, Secretaría de Posgrado,
Doctorado en Ciencia y Tecnología

Estados conformacionales ultracompactos de proteínas nativas

TESIS DOCTORAL

Leandro Grille Coronel

leandrogrille@gmail.com

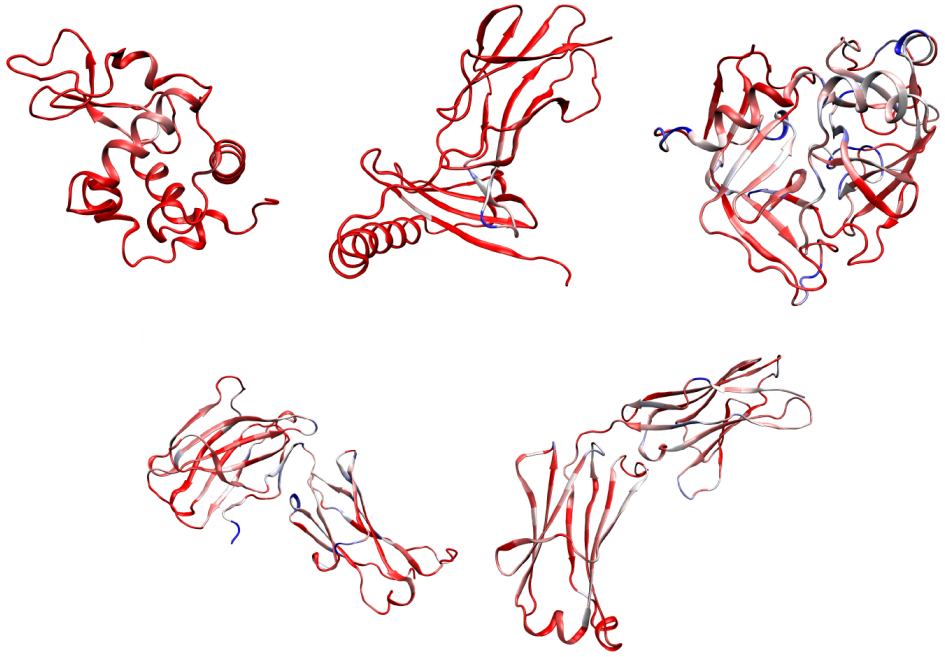
Resumen

Un análisis estadístico de alrededor de 20000 estructuras cristalográficas de proteínas puso de manifiesto efectos de la temperatura de la recolección de datos de difracción de rayos X (TCD) sobre las distancias intramoleculares y el grado de compactación. Las cadenas idénticas con datos recolectados a temperaturas ultrabajas o criogénicas (< 160 K) mostraron un radio de giro (R_G) significativamente más pequeñas que a temperaturas moderadas (> 240 K). En otro orden, el análisis reveló la existencia de estructuras con un R_G significativamente más pequeño de lo esperado para temperaturas criogénicas. En estos casos ultracompactos, el R_G inusualmente pequeño no pudo explicarse por ningún parámetro experimental específico o por características del cristal. La ultracompactación involucra a la mayoría de los átomos y resulta en su desplazamiento hacia el centro de masa (CM) de la molécula. En promedio, las estructuras ultracompactas tienen enlaces de vans der Waals y puentes de hidrógeno significativamente más cortos de lo esperado para estructuras de temperatura ultrabajas. Además, el número de contactos de vans der Waals fue mayor en las estructuras ultracompactas que en el grupo de las proteínas resueltas a temperaturas ultrabajas. Las estructuras de los estados ultracompactos fueron analizadas en detalle y en el trabajo se discuten las implicancias y las posibles causas del fenómeno.

Universidad Nacional de Quilmes

Departamento de Ciencia y Tecnología

Estados conformacionales ultracompactos de proteínas nativas



Tesis presentada para optar al título de Doctor en Ciencia y Tecnología de la Universidad Nacional de Quilmes

---

Leandro Grille Coronel

Director de Tesis: Dr. Mario R. Ermácora
Laboratorio de Expresión y Plegado de Proteínas
Bernal, 2019

Algunos de los resultados incluídos en este trabajo fueron publicados en: **Coronel, Leandro Grille, Juan P. Acierno, and Mario R. Ermácora. *"Ultracompact states of native proteins."* Biophysical chemistry 230 (2017): 36-44.**

A Carmela, Muriel y Ulises: mis hijos.

Agradecimientos

Muchas personas me ayudaron a concluir este trabajo y merecen mi sincero aprecio y reconocimiento. En primer lugar, las compañeras y compañeros que han pasado por el LEPP a lo largo y ancho de estos años: Fede, Gabi, Alejandro, Laura, Cinthia, Antonella, Raúl, Alejo, Ivana y Martín entre otras personas que han trabajado allí investigando, aprendiendo, enseñando y siempre dispuestas al aliento, a la amistad y a la ayuda.

Quiero agradecer muy especialmente a Juan Pablo Acierno, porque estuvo siempre, desde el principio y sin importar en que lugar del mundo estuviera viviendo. Sin él este trabajo simplemente no habría sido posible.

Quiero agradecer a Pamela Toledo y Ramiro Llovera. Todo lo que pueda decir de ellos es poco: como amigos, como científicos, como soporte. No tengo idea de que habría sido de mí sin ellos.

Quiero recordar a Iván que alcanzó el sueño de tocar el cielo de los Andes.

Quiero agradecer a mis amigos y amigas de ahora y de siempre. No puedo nombrar a todos sin caer en omisiones groseras e involuntarias, pero menciono a algunos en las dos riveras que son insoslayables: Nico, Manolo, Luisito, Juan Pablo, Virginia, el Chute y Milton en Uruguay, y Guido, Sonia, Gabi, Juli, José, Yamila, Conrado, Paola, Cachito, Ale y resto de los pibes del fondo, de este lado. Los quiero.

Quiero agradecer a Daniela Di Bella, mi compañera de docencia en la Universidad. Disfruté cada clase que dimos juntos, cada tren que nos tomamos, cada café que compartimos. Referente y amiga.

Quiero agradacer a todos los alumnos y alumnas que nos sufrieron. De ellos aprendí mucho más de lo que pude enseñarles.

Quiero agradecer a mi familia, a mis hijos y a Rocío. Estuvieron, están y seguirán estando como motivo primordial de todos mis actos.

Y finalmente le agradezco a Mario, que me orientó generosamente hasta el final y a la Universidad Nacional de Quilmes que me acogió y me brindó una beca que me permitió transitar este camino.

Índice general

Agradecimientos **II**

1. Introducción **4**
1.0.1. Estados y subestados conformacionales 5
1.0.2. La descripción geométrica de los polímeros mediante el Radio de Giro (R_G) 8

2. Hipótesis y objetivos **13**
2.1. Hipótesis 13
2.2. Objetivos 13

3. Materiales y métodos **16**
3.1. **Cálculo del R_G de la cadena principal** 17
3.1.1. Código C++ para cálculo de R_G 17
3.2. Normalización de los valores de R_G. Cálculo de Valor Z. 20
3.2.1. Código en R para la normalización y cálculo de Valor Z 21
3.3. Cálculo de la distancia de los átomos del esqueleto peptídico al CM 22
3.3.1. Código en C++ para la obtención de la distancias al CM de los átomos del esqueleto peptídico 22
3.4. Análisis estadístico y visualización molecular 25
3.4.1. Código en Tcl Tk de Obtención de puentes de hidrógeno 25
3.4.2. Código en Tcl Tk para obtención de contactos de van der Waals 26
3.4.3. Superposición de representaciones de Richardson 27

4. Resultados **30**
4.1. **La relación entre el R_G y la longitud de la secuencia proteica.** 30
4.2. **Contracción de las estructuras de proteínas inducida por temperatura** 33
4.3. **Estados ultracompactos de proteínas nativas** 36
4.4. Tripsina 42
4.5. β-2-microglobulina (β2M) 59
4.6. Ciclina A2 69
4.7. Quimotripsina A 72
4.8. Otros casos: lisozima, cadena pesada de Fab, cadena liviana de Fab, cadena pesada del Factor VII de coagulación y HLA DRα 77
4.8.1. Lisozima (PDB: 3IJV cadena A) 78
4.8.2. HLA-DRα (PDB: 2QW6 cadena A) 80
4.8.3. Cadena pesada de Factor de Coagulación VII (PDB: 2FLR cadena H) . . 83

4.8.4. Fragmentos Fab de anticuerpo monoclonal anti KcsA 85

5. Discusión **90**

5.1. Discusión 90

5.1.1. Compactación inducida por temperatura 92

5.1.2. Estados nativos ultracompactos 94

Índice de tablas

4.1. **Número de estructuras por rango de temperaturas y valores Z.** 33
4.2. **Estructuras ultracompactas** 37
4.3. **Estructuras ultracompactas de tripsina.** 45
4.4. **Aproximación al CM en tripsinas ultracompactas.** 47
4.5. **Estructuras ultracompactas de β2M.** 62
4.6. **Aproximación a CM en β2M ultracompactas.** 63
4.7. **Estructuras ultracompactas de Ciclinas 2A.** 70
4.8. **Aproximación a CM en ciclinas A2 ultracompactas.** 71
4.9. **Estructura ultracompacta de Quimotripsina A.** 73
4.10. **Aproximación al CM en quimotripsina A ultracompacta.** 74
4.11. **Otros casos de estructuras ultracompactas.** 78
4.12. **Aproximación a CM en casos ultracompactos (miscelánea).** 78

Índice de figuras

4.1. R_G **en función de la longitud de secuencia.** 32
4.2. **Densidad estimada de la distribución de los valores de** Z **para tres diferentes rangos de temperaturas considerados** 35
4.3. **Puentes de Hidrógeno según TCD y grado de compactación.** 40
4.4. **Contactos de van der Waals según TCD y grado de compactación.** . . 41
4.5. **Distribución de** R_G **de tripsinas y densidad de la distribución de** R_G **según TCD.** 42
4.6. **Densidad de la distribución de** R_G **en tripsinas ultrafrías.** 44
4.7. **Densidad de la distribución de** R_G **en clases de cadenas ultrafrías.** . . . 46
4.8. **Diferencias de distancias a CM de los átomos de la cadena principal en 1QL7.** 48
4.9. **Representación gráfica de la compactación con diagrama de Richardson.** 49
4.10. **Representaciones de Richardson de tripsinas ultracompactas 'superpuestas' con estructuras de referencia.** 50
4.11. **Diferencias de distancias al CM en tripsinas ultracompactas 1AQ7 Y 2A7H.** 52
4.12. **Diferencias de distancias al CM en tripsinas ultracompactas 2G81 Y 3RU4.** 54
4.13. **Estructura de benzamidinas y diamidinas aromáticas.** 55
4.14. **Diferencias de distancias al CM en tripsinas ultracompactas 3GY2, 3GY3, 3GY5, 3GY6 y 3GY8.** 57
4.15. **Gráfico de diferencia de distancia al CM.** 58
4.16. **Representaciones de Richardson de tripsinas ultracompactas superpuestas entre ellas.** 59
4.17. R_G **de tres clases de cadenas de** β**-2-microglobulina (β2M).** 61
4.18. **Diferencias de distancias al CM en β2M de PDB: 3VFN.** 64
4.19. **Diferencias de distancias al CM en β2M de PDB: 1T1Z.** 65
4.20. **Diferencias de distancias al CM en las β2M de PDB: 3UTT.** 66
4.21. **Representaciones de Richardson de β2M humanas ultracompactas 'superpuestas' con estructuras de referencia.** 67
4.22. **Diferencias de distancias al CM en β2M murina de PDB: 1K8D.** 68
4.23. **Representación de Richardson de β2M murina ultracompacta (PDB: 1K8D) superpuesta con β2M de referencia (PDB: 3P4N).** 68
4.24. **Gráfico de dispersión de** R_G **en la clase de Ciclina A2.** 69
4.25. **Diferencias de distancias al CM en las ciclinas 2WMA-B y 2WMA-D.** 71

4.26. **Representaciones de Richardson de ciclinas A2 humanas ultracompactas superpuestas con la estructura de referencia 1QMZ-B.** 72
4.27. **Gráfico de dispersión de R_G en la clase de la quimotripsina A.** 73
4.28. **Diferencias de distancias al CM en la quimotripsina de PDB: 3RU4-D.** 75
4.29. **Representación de Richardson de la quimotripsina ultracompacta 3RU4-D superpuesta con estructura de referencia 1YPH-D.** 76
4.30. **Representaciones de Richardson de cadenas ultracompactas superpuestas con estructuras de referencia de sus clases respectivas.** 77
4.31. **Gráfico de dispersión de R_G en la clase de las lisozimas resueltas a temperatura ultrabaja.** 79
4.32. **Diferencias de distancias al CM en la lisozima 3JIV-A.** 80
4.33. **Gráfico de dispersión de los R_G en la clase de las HLA DRα resueltas a temperatura ultrabaja.** 81
4.34. **Diferencias de distancias al CM en la HLA DRα 2Q6W-A.** 83
4.35. **Gráfico de dispersión de los R_G de los factores de coagulación VII resueltos a temperatura ultrabaja.** 84
4.36. **Diferencias de distancias al CM en el factor de coagulación VII de PDB: 2FLR-H.** 85
4.37. **Gráficos de dispersión de las clases del FAb anti KscA: cadena liviana y pesada.** 86
4.38. **Diferencias de distancias al CM de los átomos de las cadenas principales de 2DWD-B y 2DWD-A.** 87

Abreviaturas

β2M Beta-2-microglobulina.

BCTI Inhibidores Bowman Birk de serina proteasas.

CDK Quinasa dependiente de Ciclina.

CM centro de masa de proteína.

CMH Complejo Mayor de Histocompatibilidad.

HLA Antígeno Leucocitario Humano.

InterPro Protein sequence analysis and classification of EMBL.

PDB Protein Data Bank.

R_G radio de giro.

TCD temperatura de recolección de datos de difracción.

UniProtKB The UniProt Knowledgebase.

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

Introducción

Capítulo 1

Introducción

En el presente trabajo investigamos la existencia de subestados conformacionales compactos y ultracompactos de proteínas resueltas por difracción de Rayos X.

Denominamos "subestado conformacional compacto" de una proteína nativa a aquél que manifiesta una contracción hacia el centro de masa CM de la estructura, resultando en un encogimiento mensurable de carácter global sin que se alteren las longitudes de los enlaces covalentes de su esqueleto polipeptídico. Todo ello en comparación con la estructura promedio de las proteínas de idéntica secuencia.

Cabe destacar que una proteína compacta, en el sentido que se define en este trabajo, no es una estructura nativa que ha sido sometida a altas presiones o algún tipo de fuerza mecánica contractiva que reduzca su volumen. Es una proteína que ha alcanzado la estructura tridimensional nativa en condiciones idénticas a las de su clase, pero con mayor densidad de empaquetamiento y, por lo tanto, en un menor volumen.

Conviene introducir algunos conceptos de la teoría del plegado de proteínas, revisar antecedentes y discutir los parámetros físicos utilizados para desarrollar la investigación.

1.0.1. Estados y subestados conformacionales

Las proteínas son macromoléculas biológicas que alcanzan su conformación nativa en condiciones fisiológicas. De acuerdo a la teoría del plegado de proteínas, el estado nativo es la estructura tridimensional más estable y se encuentra en el mínimo de la hipersuperficie de energía libre de Gibbs en función de las coordenadas conformacionales [49]. Durante mucho tiempo se creyó que el estado nativo era único en contraste con los estados desplegados, que se distribuyen como una población estadística. Sin embargo, esta creencia largamente sostenida, posiblemente tributaria de que la técnica dominante en los estudios estructurales ha sido la cristalografía de rayos X [23], se ha probado inexacta: el estado nativo no es estrictamente único. De la biología de las enzimas se desprende que muchas proteínas son nanomáquinas que admiten por lo menos un estado abierto y un estado cerrado [20], y deben fluctuar entre estos estados para permitir el ingreso de su ligando específico, pero además cada estado se distribuye en una estructura fina de subestados conformacionales porque los valles de energía son rugosos y, por lo tanto, cabe afirmar que hay valles dentro de los valles y valles dentro de los valles que están dentro de los valles, en una bifurcación sucesiva que refleja la degeneración local de la superficie de energía [1, 20, 23, 21].

Los subestados conformacionales son conformaciones con variaciones sutiles entre ellas y esencialmente isoenergéticas que se alcancanzan a través de fluctuaciones en el equilibrio mientras que el tránsito entre un estado conformacional y otro involucra movimientos importantes para la función que se producen fuera del equilibrio [1, 20]. Los subestados conformacionales no tienen una repercusión mayor en la especificidad de la función, pero pueden afectar la velocidad de los procesos químicos que protagonizan y una de las primeras evidencias de su existencia proviene del estudio de la dinámica de reasociación entre la mioglobina y sus ligandos CO u O_2

a baja temperatura, luego de la aplicación de un pulso de láser a una longitud de onda precisa para romper el enlace convalente mediante la técnica de fotólisis de flash [3, 20]. A temperatura ambiente el CO libre difunde al solvente, pero a temperaturas ultrabajas o criogénicas el CO queda atrapado en el bolsillo del hemo y tiende a reenlazarse con una cinética que la teoría prevé exponencial con el tiempo, habida cuenta la superficie potencial de doble pozo (un mínimo donde el hierro hemo está unido convalentemente al ligando y un mínimo que corresponde a la conformación de la proteína con el hierro libre) separado por una barrera de energía. Sin embargo, el experimento demuestra que la reasociación se aleja de la cinética exponencial prevista y asume un comportamiento potencial que sólo puede ser explicado por la presencia de subestados conformacionales con variaciones pequeñas de la afinidad por el ligando [3, 20, 23, 21].

Los subestados conformacionales conservan la estructura primaria, secundaria y terciaria del estado conformacional del que derivan, pero varían en detalles estructurales finos y se corresponden con mínimos locales separados por pequeñas barreras de energía en el contexto de un mínimo global de la superficie de energía [20, 23, 21]. En condiciones fisiológicas o a temperaturas lo suficientemente altas, las proteínas fluctúan entre los subestados conformacionales de forma continua, pero a temperaturas por debajo de cierto umbral (180 K), ya no pueden superar las barreras y quedan atrapadas en subestados con propiedades físicas ligeramente distintas entre ellos. Fraunfelder denomina este último fenómeno de "ruptura de la simetría de subestados" [20, 23, 21].

La hipótesis de la existencia de subestados conformacionales ha sido evaluada por diversas técnicas y por muchos otros grupos de investigadores, llegando a resultados consistentes y concluyentes [20, 3, 24, 27] . Los subestados conformacionales están organizados jerárquicamente y

pueden ser categorizados en subestados "taxonómicos", "estadísticos" o "locales" [20, 21, 23]. Un subestado taxonómico o isómero es una conformación que puede ser caracterizada individualmente y su estructura se puede determinar de forma aislada. Corresponden al primer nivel de los subestados, pero como se dijo antes, en cada subestado las proteínas pueden adoptar conformaciones ligeramente distintas, un segundo nivel de degeneración, explorando valles que componen las estructura fina de ese estado basal. Los subestados estadísticos ya no pueden ser descritos por sí mismos, sino como ensamble y sus existencias se revelaron por técnicas experimentales que indagaban sobre la cinética de reasociación de un subestado taxonómico al ligando en condiciones criogénicas, observando el perfil no exponencial de la reacción de reenlance en función del tiempo, propio de la presencia de una multitud de isoformas con ritmos de asociación distintos [20, 21, 23]. En un tercer nivel, los subestados locales son pequeñas variaciones en la conformación de los subestados estadísticos cuya existencia ha podido ser relevada [20, 21], pero todo indica que es posible descender todavía más dentro de las rugosidades de los mínimos de la hipersuperficie energía y que la cantidad total de subestados de un estado conformacional, pese a no haber sido determinado, debe exceder el número de 10^{10} [20, 21, 23].

La existencia de subestados conformacionales también ha sido detectada por difracción de rayos X [24, 27] . Aunque durante mucho tiempo se consideró que los datos obtenidos por la técnica de cristalografía de rayos X justamente probaban lo contrario: que las moléculas de proteínas en estado nativo tenían una conformación única. Sin embargo, un análisis más fino de los espectros de difracción revelaron que el comportamiento de un cristal de proteína era distinto al de un sólido cristalino ideal, lo que analizado a la luz de los fundamentos téoricos de esta técnica experimental aportaba evidencias de que, en el caso de las proteínas, el patrón difractivo corresponde a moléculas con una estructura global conservada, pero con dismilitudes puntuales o pequeñas variaciones confguracionales. Básicamente, en un sólido las posiciones equivalentes

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

en la red cristalina están relacionadas por simetría y la desviación media cuadrática de la posición de los átomos es isotrópica y muy pequeña: del mismo orden de magnitud para todos los átomos. Por el contrario, en el caso del cristal de una molécula de proteínas, la existencia de subestados conformacionales determina que haya átomos que no están en la misma posición de equilibro, y por lo tanto la desviación media cuadrática del conjunto es mucho más grande que para un sólido cristalino y, además, no es uniforme. Esto es: la existencia de un ensamble de moléculas con una estructura general casi, pero no exactamente idéntica, contribuye de forma significativa al factor de Debye-Waller (factor B) que ya no es una variable unícamente dependiente de la vibración térmica alrededor de la posición de equilibrio, inevitable incluso a temperaturas ultrabajas para cualquier molécula cristalina, sino también de la existencia de subestados con matices conformacionales [20, 21, 23].

1.0.2. La descripción geométrica de los polímeros mediante el Radio de Giro (R_G)

Un parámetro morfológico sensible a cambios globales en la estructura tridimensional de las moléculas de proteínas es el R_G [55]. En el campo de la biología estructural, R_G es una propiedad útil debido a su significado físico directo, y a que puede ser medido de forma experimental, pero también estimado mediante cálculos [29]. Matemáticamente, R_G se define como el valor medio cuadrático de la distancia de las partes de un objeto a su centro de masa. En consecuencia, en una macromolécula puede ser computado si se conoce las coordenadas espaciales y la identidad química de los átomos que la componen. El R_G es una propiedad física característica de la proteína en su estado nativo, como lo son sus dominios de plegado, la estructura terciaria o el número y posición de hélices alfa y hebras beta. Asímismo, por su definición, las tres componentes ortogonales del parámetro son medidas de tamaño y, conjun-

tamente, proporcionan información sobre la forma, por lo que puede ser utilizado para evaluar la expansión o contracción de proteínas en condiciones específicas [22].

La mayoría de las proteínas se pliegan en dominios globulares, que se caracterizan por alcanzar una compactación óptima y exponer una superficie mínima al solvente. En las estructuras cristalinas depositadas en el Protein Data Bank (PDB), los valores de R_G, como medida de la disposición de los átomos en el espacio físico e indicador apropiado del estado de compactación de la estructura [38], se distribuyen de forma muy estrecha alrededor de un valor promedio para todas las estructuras depositadas de una misma proteína, pese a diferencias en las condiciones experimentales en las que fueron caracterizadas. Esta distribución de los valores de R_G con muy escasa dispersión entre las estructuras cristalizadas de una misma proteína indican que el tamaño y el grado de compactación de los estados nativos se conservan aún cuando las condiciones experimentales de cristalización y resolución espacial por dispersión de rayos X varíen ostensiblemente.

Desde los albores de la cristalografía de proteínas muchos grupos han investigado la propiedades estructurales de la proteínas cristalizadas, incluyendo la sensibilidad de parámetros físicos como el R_G ante variaciones en la condiciones experimentales. En ese contexto, algunas investigaciones han aportado información sobre el comportamiento térmico del volumen y el R_G de estructuras cristalinas mediante el análisis de los patrones de difracción de los rayos X a diferentes temperaturas.

Tan temprano como en 1982, Walter et al. reportaron que las estructuras de tripsinógeno que habían determinado a 103 K y a 173 K parecían expandirse a mayor temperatura [72]. De forma contemporánea, Hartmann y Freuenfelder trabajaron con la mioglobina de cachalote y analizaron los datos cristalográficas con los patrones de difracción colectados a 80 K y en el

intervalo de 225 a 300 K, registrando un incremento del volumen de la estructura cristalina de la proteína hasta 3 % debido al calentamiento.

En dicho trabajo, los investigadores observaron una expansión térmica anisotrópica que se produce entre los 80 K y la temperatura ambiente, concentrándose los cambios mayores en el segmento que separa las hélices C y D y el segmento de proteína que separa las hélices G y H (conocidos como esquina CD y esquina GH, respectivamente) de la mioglobina de cachalote [27, 22].

Un poco más tarde, Earnest et al. [15] compararon la estructura de tripsina de rata resuelta por cristalografía a baja temperatura (LT, 120 K) con la estructura resuelta a temperatura ambiente (RT, 295 K), y se encontraron con una reducción significativa en las dimensiones de la celda unitaria cuando iban de temperatura ambiente a bajas temperatuas, que era acompañado por un decremento en el área de la superfice molecular y en el R_G .

Para estudiar el efecto de la temperatura en la estructura cristalina de la ribonucleasa A, Tilton y sus colaboradores la caracterizaron por cristalografía de rayos X, realizando la colecta de los datos de difracción a nueve temperaturas distintas en el rango de 98 K a 320 K. El análisis demostró que la estructura cristalina de la proteína experimentaba una expansión ligera (aproximadamente 0.4 % por cada 100 K) a temperatura creciente, y determinaron que dicha expansión tenía un comportamiento lineal [68].

Otros grupos estudiaron la elasticidad de R_G frente a otras variables experimentales. En particular, algunos trabajos han demostrado variaciones significativas en el R_G de proteínas unidas a ligandos, tanto en solución [54, 36, 46, 64, 47, 48, 28, 18] como en complejos cristalizados [5]. No obstante, en contraste con la expansión térmica precitada, estos cambios conformacionales, inducidos por la interacción con el substrato, usualmente consisten en cie-

rres de las hendiduras de unión a ligando debido a movimientos de pivot y rotación de lóbulos separados, por lo que deben ser clasificados como movimientos locales de tipo bisagra en lugar de una de compactación de la proteína propiamente dicha.

Un antecedente cercano y reciente se halla los resultados de Risso et al. en nuestro laboratorio. En su investigación produjo mediante ingeniería genética una forma nativa de β-lactamasa de Clase A, cuyo compactación era mayor que el resto de las β-lactamasas de su clase depositadas previamente en el PDB. En este trabajo [61] se reportó una forma mutante de la ESP (penicilinasa exo-small de *Bacillus licheniformis*) en la que se sustituye un residuo de triptofano, presente en la forma salvaje, por una fenilalanina. La mutante fue caracterizada por cristalografía de rayos X, alcanzándose una resolución de 1.7 Å de estructura mediante cristalografía de rayos X. También fueron determinados sus parámetros de actividad catalítica y su estabilidad termodinámica. De la caracterización surgió que la mutante no mostró cambios cualitativos de conformación comparado con las 13 estructuras de β-lactamasas de *B. licheniformis* previamente reportadas, pero sí un R_G significativamente reducido. En efecto, la mutante mostró el menor R_G entre las 126 β-lactamasas de Clase A depositadas en el PDB.

Los antecedentes reseñados en esta Introducción sugieren que el mero análisis del R_G , un parámetro simple de determinar y directamente relacionado con la densidad del empaquetamiento de la estructuras cristalinas, es por sí mismo capaz de proporcionar información estructural valiosa para el análisis biofísico de las proteínas y su plegado.

Hipótesis y objetivos

Capítulo 2

Hipótesis y objetivos

2.1. Hipótesis

(a) Las proteínas nativas en general manifiestan una compactación térmica en el rango de temperaturas utilizadas en la recolección de datos de difracción de rayos X.

(b) La compactación de las proteínas en el estado cristalino no es función sencilla de la temperatura sino un fenómeno histerético con distintas manifestaciones de acuerdo a las condiciones experimentales.

(c) Existen conformaciones ultracompactas de las proteínas cristalinas que no pueden explicarse simplemente por el efecto de la temperatura.

2.2. Objetivos

(a) Desarrollar métodos robustos para cuantificar el grado de compactación de proteínas cristalinas.

(b) Utilizar los métodos desarrollados para el análisis estadístico de los estructuras cristalinas depositadas en el Protein Data Bank (PDB).

(c) Identificar casos de proteínas compactas y ultracompactas en las distintas familias de proteínas representadas en el PDB.

(d) Formular nuevas hipótesis para explicar el comportamiento térmico de las estructuras proteicas cristalinas en el marco de las teorías modernas del plegado proteico.

Materiales y métodos

Capítulo 3

Materiales y métodos

Una lista inicial de códigos de Identificación (ID) de cadenas de proteínas fue descargada del sitio PISCES (pdbaanr; 2015). La lista incluye 65,195 clases de cadenas. Cada clase de cadena incluye varios resoluciones experimentales de la misma secuencia en una o más entradas del PDB. El número total de cadenas el conjunto inicial fue de 249,185 en total, correspondientes a 95,503 entradas del PDB. Por ejemplo, la clase de cadenas de la hemoglobina W37A (con mutación de triptofano a alanina en posición 37) contiente cuatro cadenas: 1A01 B, 1A01 D, 1A0W B Y 1A0W D, de dos entradas del PDB distinas (la 1A01 y la 1A0W). Las cuatro cadenas de hemoglobina tienen la misma identidad de secuencia y sus estructuras resueltas deben considerarse réplicas experimentales de la misma cadena pero en diferentes condiciones.

El set inicial fue tratado por medio de una serie de pasos de depuración: (i) sólo se consideraron cadenas con una resolución superior a 3.0 Å; (ii) se descartaron las cadenas con átomos faltantes, inconsistencias geométricas u otras anomalías experimentales, y (iii) las clases de cadenas con menos de 20 representantes fueron eliminadas.

El conjunto final analizado en el trabajo consistió en 631 clases de cadenas, cada una de ellas con por lo menos 20 réplicas de secuencia idéntica. Considerando todas las clases, el espacio muestral abarcó 19,393 cadenas de 7,114 entradas distintas del PDB.

3.1. Cálculo del R_G de la cadena principal

El R_G de la cadena principal describe bien la forma y el tamaño de la molécula mediante el cómputo de la dispersión de los átomos individuales del esqueleto peptídico alrededor del CM o el centro geométrico. En nuestro trabajo, este R_G fue estimado utilizando el centro geométrico, dado que en proteínas puede asumirse la identidad de masas en todos los átomos distintos del hidrógeno y, en esa aproximación, el CM coincide con el centroide. Más aún: si este supuesto de equivalencia general fuera inadecuado, no lo es en el esqueleto de las proteínas, por cuanto la distribución de los átomos de la cadena principal es periódica y homogénea y, en consecuencia, en el cálculo todas las masas se cancelan. El vector del centro geométrico ($\mathbf{R}_c$) se define:

$\sum_i^N (\mathbf{r}_i - \mathbf{R}_c) = 0$

donde $\mathbf{r}_i$ es el vector coordenadas del átomo iésimo de la cadena principal, de donde:

$\mathbf{R}_c = \sum_i^N (\mathbf{r}_i)/N$

siendo N el número total de átomos considerados.

A partir del centro geométrico definido, R_G es un escalar que se aproxima como:

$R_g \simeq \sqrt{\sum_i^N (\mathbf{r}_i - \mathbf{R}_c)^2/N}$

3.1.1. Código C++ para cálculo de R_G

```
#include <vector>
#include <cstring>
#include <cstdlib>
#include <string>
#include <fstream>
#include <iostream>
#include <cmath>
#include <iomanip>
using namespace std;
```

```
10
11 int main()
12 {
13   //cout << "Escriba el nombre del archivo a leer: " << endl;
14   string filename;
15   cin >> filename;
16   //cout << filename<<"";
17   ifstream file(filename.c_str());
18   if(!file)
19   {
20     cout << "Error no se puede abrir el archivo: " << filename << endl;
21                 return 1;
22   }
23
24 string linea, archivox, archivoy, archivoz;
25 /*float coord [num_col] [num_fil];*/
26
27 float x=0, y=0, z=0;
28 typedef std::vector<float> Array;
29 typedef std::vector<Array> TwoDArray;
30 int  numRows=0,numCols=3;
31 string C=(" C  "), CA=(" CA "), N=(" N  "), O=(" O  ");
32 TwoDArray Arr2D;
33
34
35
36 while (getline(file, linea))  {
37
38  if ((linea.compare(12,4," C  ")==0)
39 || (linea.compare (12,4, " CA ")==0)
40 || (linea.compare(12,4,  " N  ")==0)
41 || (linea.compare(12,4,  " O  ")==0))
42 {
43         // Fill in test data
44       archivox = linea.substr(30,8);
45         x = atof (archivox.c_str());
46         archivoy = linea.substr(38,8);
47         y = atof (archivoy.c_str());
48         archivoz = linea.substr(46,8);
49         z = atof (archivoz.c_str());
50
51     Arr2D.push_back(Array());
52     for (int j = 0; j < numCols; j++) {
53         Arr2D[numRows].push_back(x); //aca va el atof(x)
54         Arr2D[numRows].push_back(y); //aca va el atof(y)
55         Arr2D[numRows].push_back(z); //aca va el atof(z)
56     }
57
58     /* contador para conocer rl numRows*/
59     numRows++;
60 }}
61
62 // Make sure the data is there
63 /*for (int i = 0; i < numRows; i++) {
64     for (int j = 0; j < numCols; j++) {
```

```
                std::cout << Arr2D[i][j] << "\t\t";
        }
cout << '';*/
}
float sumxyz [3]={0,0,0};
for (int i=0; i< numRows; i++)
    {
sumxyz [0]+=Arr2D[i][0];
sumxyz [1]+=Arr2D[i][1];
sumxyz [2]+=Arr2D[i][2];}
float cmxyz [3]={0,0,0};
cmxyz [0]=sumxyz[0]/numRows;
cmxyz [1]=sumxyz[1]/numRows;
cmxyz [2]=sumxyz[2]/numRows;

//cout << "Centro de Masa ="<< cmxyz [0] <<"\t"<< cmxyz[1]<< "\t"<<cmxyz [2]<<
    endl;

float RGxyz [3]={0,0,0};
float RG;
for (int i=0; i< numRows; i++)
    {
RGxyz [0] += (Arr2D[i][0] -cmxyz [0]) *(Arr2D[i][0] -cmxyz [0]);
RGxyz [1] += (Arr2D[i][1] -cmxyz [1]) *(Arr2D[i][1] -cmxyz [1]);
RGxyz [2] += (Arr2D[i][2] -cmxyz [2]) *(Arr2D[i][2] -cmxyz [2]);}

float RGxyz_N [3]={0,0,0};
float RG_N;
for (int i=0; i<numRows;i+=4)
{
RGxyz_N [0] += (Arr2D[i][0] -cmxyz [0]) *(Arr2D[i][0] -cmxyz [0]);
RGxyz_N [1] += (Arr2D[i][1] -cmxyz [1]) *(Arr2D[i][1] -cmxyz [1]);
RGxyz_N [2] += (Arr2D[i][2] -cmxyz [2]) *(Arr2D[i][2] -cmxyz [2]);}

float RGxyz_CA [3]={0,0,0};
float RG_CA;
for (int i=1; i<numRows;i+=4)
{
RGxyz_CA [0] += (Arr2D[i][0] -cmxyz [0]) *(Arr2D[i][0] -cmxyz [0]);
RGxyz_CA [1] += (Arr2D[i][1] -cmxyz [1]) *(Arr2D[i][1] -cmxyz [1]);
RGxyz_CA [2] += (Arr2D[i][2] -cmxyz [2]) *(Arr2D[i][2] -cmxyz [2]);}

float RGxyz_C [3]={0,0,0};
float RG_C;
for (int i=2; i<numRows;i+=4)
{
RGxyz_C [0] += (Arr2D[i][0] -cmxyz [0]) *(Arr2D[i][0] -cmxyz [0]);
RGxyz_C [1] += (Arr2D[i][1] -cmxyz [1]) *(Arr2D[i][1] -cmxyz [1]);
RGxyz_C [2] += (Arr2D[i][2] -cmxyz [2]) *(Arr2D[i][2] -cmxyz [2]);}

float RGxyz_O [3]={0,0,0};
float RG_O;
for (int i=3; i<numRows;i+=4)
{
RGxyz_O [0] += (Arr2D[i][0] -cmxyz [0]) *(Arr2D[i][0] -cmxyz [0]);
```

```
RGxyz_O [1] += (Arr2D[i][1] -cmxyz [1]) *(Arr2D[i][1] -cmxyz [1]);
RGxyz_O [2] += (Arr2D[i][2] -cmxyz [2]) *(Arr2D[i][2] -cmxyz [2]);}

float SUMDIST_CN [3]={0,0,0};
float DIST_CN;
for (int i=2; i<numRows-2;i+=4)

{
SUMDIST_CN [0] += (Arr2D[i][0] -Arr2D[i+2][0]) *(Arr2D[i][0] -Arr2D[i+2][0]);
SUMDIST_CN [1] += (Arr2D[i][1] -Arr2D[i+2][1]) *(Arr2D[i][1] -Arr2D[i+2][1]);
SUMDIST_CN [2] += (Arr2D[i][2] -Arr2D[i+2][2]) *(Arr2D[i][2] -Arr2D[i+2][2]);}


DIST_CN = sqrt((float) (SUMDIST_CN [0] + SUMDIST_CN [1] + SUMDIST_CN [2]))
/ sqrt ((float) (numRows/4));

RG = sqrt ((float) (RGxyz [0] + RGxyz [1] + RGxyz [2]))
/ sqrt ((float) numRows);
RG_N=sqrt ((float) (RGxyz_N [0] + RGxyz_N [1] + RGxyz_N [2]))
/ sqrt ((float) (numRows/4));
RG_CA=sqrt ((float) (RGxyz_CA [0] + RGxyz_CA [1] + RGxyz_CA [2]))
/ sqrt ((float) (numRows/4));
RG_C=sqrt ((float) (RGxyz_C [0] + RGxyz_C [1] + RGxyz_C [2]))
/ sqrt ((float) (numRows/4));
RG_O=sqrt ((float) (RGxyz_O [0] + RGxyz_O [1] + RGxyz_O [2]))
/ sqrt ((float) (numRows/4));

cout<< setw(8)<<filename <<"\t"<<setw(5)<<numRows<<"\t"<<setw(8)
<<setprecision(8)<<RG <<"\t"<<setw(5)<<numRows/4
<<"\t"<<setw(8)<<setprecision(8)<<RG_N<<"\t"<<setw(5)
<<numRows/4<<"\t"<<setw(8) <<setprecision(8)<<RG_CA<<"\t"<<setw(5)
<<numRows/4<<"\t"<<setw(8)<<setprecision(8)<<RG_C<<"\t"<<setw(5)
<<numRows/4<<"\t"<<setw(8)<<setprecision(8)<<RG_O<<"\t"<<setw(5)
<<numRows/4<<"\t"<<setw(8)<<setprecision(8)<<DIST_CN<<"";

return 0;
}
```

3.2. Normalización de los valores de R_G. Cálculo de Valor Z.

Los valores de R_G fueron normalizados de forma de poder comparar cadenas de diferentes clases de cadenas. Para ello, se obtuvo el valor de la variable centrada reducida que consiste en sustraer la media a cada uno de los valores de R_G y dividirlos por la desviación típica de la variable en su clase:

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

$Z_i = (R_{g,i} - \bar{R_{g,j}})/\mathbf{S}_{R_{g,j}}$

donde $R_{g,i}$ es el R_G de la cadena i perteneciente a la clase de cadenas j; $\bar{R_{g,j}}$ es el promedio de los R_G en la clase de cadenas j; y $S_{R_{g,j}}$ es la deviaciòn estándar de R_G dentro su clase j de pertenencia.

3.2.1. Código en R para la normalización y cálculo de Valor Z

```


MyFileList<-scan(file="lista_de_archivos_RG_de_clase.txt",
list(filename=""), sep="",skip=0,nlines=0)
nfiles<-length(MyFileList$filename)
for(i in 1:nfiles) {
mydata<-read.table(file=MyFileList$filename[i],
fill=TRUE, sep="\t",na.strings="NULL")

attach(mydata)

temp<-table (as.vector(V2))

length(as.vector(V2))->registros

mcv<-names(temp)[temp == max(temp)]

Mmcv<-max(mcv)

print(MyFileList$filename[i])

which(V2 !=Mmcv | V4>1.35| V4<1.29)->index

mydata[index,]->registros_excluidos

length(registros_excluidos$V2)->reg_ex

print(reg_ex)

detach()

if (length(index)!=0 & registros-reg_ex>=5){

mydata1<-mydata[-index,]

attach(mydata1)

mydata1$V6<-(V3-mean(V3))/(sd(V3))

write.table(mydata1, file=paste (MyFileList$filename[i],
 "edt", sep="."), sep="\t", row.names=FALSE,
  col.names=FALSE, na="NULL")
```

```
43
44 detach()}
45
46 if (length(index)==0 & registros>=5){
47
48 mydata2<-mydata
49
50 attach(mydata2)
51
52 mydata2$V6<-(V3-mean(V3))/(sd(V3))
53
54 dataout<-paste (MyFileList$filename[i], "prob", sep=".")
55
56 for(i in 1:length(mydata2)){
57
58 sink(file = dataout, append = FALSE, type = "output", split = FALSE)
59
60   cat(mydata2$V1[i],mydata$V2[i],mydata$V3[i],mydata2$V4[i],
61   mydata2$V5[i],mydata2$V6[i],"",file="", sep="\t")
62 }
63
64 detach()}
65 sink()
66 }
```

3.3. Cálculo de la distancia de los átomos del esqueleto peptídico al CM

La distancia al CM de los átomos del esqueleto peptídico de las proteínas fue calculado a partir de las coordenadas espaciales de cada átomo según se registran en el depósito del PDB, y utilizando las coordenadas del CM obtenidas en la determinación de los R_G.

3.3.1. Código en C++ para la obtención de la distancias al CM de los átomos del esqueleto peptídico

```
1 #include <vector>
2 #include <cstring>
3 #include <cstdlib>
4 #include <string>
5 #include <fstream>
6 #include <iostream>
7 #include <cmath>
8 #include <iomanip>
9 using namespace std;
10
```

```
11 int main()
12 {
13   //cout << "Escriba el nombre del archivo a leer: " << endl;
14   string filename;
15   cin >> filename;
16   cout << filename<<"\t";
17   ifstream file(filename.c_str());
18   if(!file)
19   {
20     cout << "Error no se puede abrir el archivo: " << filename << endl;
21                 return 1;
22   }
23
24 string linea, archivox, archivoy, archivoz;
25 /*float coord [num_col] [num_fil];*/
26
27 float x=0, y=0, z=0;
28 typedef std::vector<float> Array;
29 typedef std::vector<Array> TwoDArray;
30 int  numRows=0,numCols=3;
31 string C(" C  "), CA(" CA "), N=(" N  "), O=(" O  ");
32 TwoDArray Arr2D;
33
34
35
36 while (getline(file, linea))  {
37
38  if ((linea.compare(12,4," C  ")==0)
39 || (linea.compare (12,4, " CA ")==0)
40 || (linea.compare(12,4,  " N  ")==0)
41 || (linea.compare(12,4,  " O  ")==0))
42 {
43         // Fill in test data
44       archivox = linea.substr(30,8);
45         x = atof (archivox.c_str());
46         archivoy = linea.substr(38,8);
47         y = atof (archivoy.c_str());
48         archivoz = linea.substr(46,8);
49         z = atof (archivoz.c_str());
50
51     Arr2D.push_back(Array());
52     for (int j = 0; j < numCols; j++) {
53         Arr2D[numRows].push_back(x); //aca va el atof(x)
54         Arr2D[numRows].push_back(y); //aca va el atof(y)
55         Arr2D[numRows].push_back(z); //aca va el atof(z)
56     }
57
58     /* contador para conocer rl numRows*/
59     numRows++;
60 }}
61
62 // Make sure the data is there
63 for (int i = 0; i < numRows; i++) {
64     for (int j = 0; j < numCols; j++) {
65             //std::cout << Arr2D[i][j] << "\t\t";
```

```
66      }
67 //cout << '';
68 }
69 float sumxyz [3]={0,0,0};
70
71 for (int i=0; i< numRows; i++)
72 {
73 sumxyz [0]+=Arr2D[i][0];
74 sumxyz [1]+=Arr2D[i][1];
75 sumxyz [2]+=Arr2D[i][2];}
76
77 float cmxyz [3]={0,0,0};
78 cmxyz [0]=sumxyz[0]/numRows;
79 cmxyz [1]=sumxyz[1]/numRows;
80 cmxyz [2]=sumxyz[2]/numRows;
81
82 //cout << "Centro de Masa ="<< cmxyz [0] <<"\t"<< cmxyz[1]<< "\t"<<cmxyz [2]<<
     endl;
83
84 float DIST_CM_xyz [3] ={0,0,0};
85 float DIST_\gls{cm};
86 for (int i=0; i< numRows; i++)
87    {
88 DIST_CM_xyz [0] = (Arr2D[i][0] -cmxyz [0]) *(Arr2D[i][0] -cmxyz [0]);
89 DIST_CM_xyz [1] = (Arr2D[i][1] -cmxyz [1]) *(Arr2D[i][1] -cmxyz [1]);
90 DIST_CM_xyz [2] = (Arr2D[i][2] -cmxyz [2]) *(Arr2D[i][2] -cmxyz [2]);
91 DIST_CM = sqrt ((float) (DIST_\gls{cm}_xyz [0] + DIST_\gls{cm}_xyz [1] + DIST_CM
     _xyz [2]));
92
93
94
95 if (i%4==0)
96 cout /*<<"N"<<"\t"*/<<DIST_\gls{cm}<<"\t";
97 else if ((i-1)%4==0)
98 cout /*<<"CA"<<"\t"*/<<DIST_\gls{cm}<<"\t";
99 else if ((i-2)%4==0)
100 cout /*<<"C"<<"\t"*/<<DIST_\gls{cm}<<"\t";
101 else if ((i-3)%4==0)
102 cout /*<<"O"<<"\t"*/<<DIST_\gls{cm}<<"\t";
103
104 }
105
106 cout<<endl;
107 return 0;
108 }
```

3.4. Análisis estadístico y visualización molecular

Los análisis estadísticos fueron realizados usando R [56]. Para evaluar la significación estadística de las diferencias entre las medias de R_G, longitud y número de puentes de de hidrógeno, cantidad y distancia promedio de contactos de van de Waals, en diferentes condiciones de temperatura o grado de compactación, se utilizaron, según los casos, la prueba T de Welch para muestras independentes y la prueba no-paramétrica U de Mann Whitney. La homogeneidad de la varianza (homocedasticidad) fue investigada con prueba de Levene. Los cálculos interatómicos y la visualización molecular se realizaron utilizando VMD [30]. La medición y de cuantificación de puentes de hidrógeno en estructuras nativas se efectuaron considerando una distancia de corte de 4.0 Å entre aceptor — donor y 40 grados de desviación máxima de $180^{\underline{o}}$ del ángulo formado por el átomo donor, hidrógeno y átomo aceptor. En el análisis de los contactos de van der Waals, sólo se consideraron los átomos de carbono de residuos diferentes y una distancia de corte máxima de 4.0 Å y mímina de 3.4 Å, correspondiente a la zona de exclusión debido a la suma de los radio de van der Waals de los átomos interactuantes.

3.4.1. Código en Tcl Tk de Obtención de puentes de hidrógeno

```
foreach pdb [lsort [glob *.pdb]] {
mol new $pdb
set pdb_e [atomselect top "protein"]
$pdb_e writepdb ${pdb}_e.pdb }
mol delete all
foreach pdb_edited [lsort [glob *_e.pdb]] {
mol new $pdb_edited
play renumerar.tcl
package require psfgen
topology top_all27_prot_lipid.inp
pdbalias residue HIS HSE
pdbalias atom ILE CD1 CD
segment U {pdb ${pdb_edited}.pdb}
coordpdb ${pdb_edited}.pdb U
guesscoord
writepdb ${pdb_edited}_1.pdb
writepsf $pdb_edited}_1.psf
}
```

```
19 mol delete all
20 foreach pdb_conteo[lsort [glob *_1.pdb]] {
21 set chan [open ${pdb_conteo}out.dat w]
22 set chan2 [open ${pdb_conteo}hb.log.dat w]
23
24 set molname ${pdb_conteo}
25
26   set mysel [atomselect top "protein and element N O"]
27   set mylist [measure hbonds 4.0 40 $mysel]
28   set mydonorsindexes [lindex $mylist 0]
29   set myacceptorsindexes [lindex $mylist 1]
30   set myhydrogensindexes [lindex $mylist 2]
31
32   set sum 0
33   set count 0
34
35   puts $chan "molname\tres1\tname1\tindex1\tres2\tname2\tindex2\tdis"
36   foreach donor $mydonorsindexes acceptor $myacceptorsindexes {
37   set myseldonors [atomselect top "index $donor"]
38   set myselacceptors [atomselect top "index $acceptor"]
39   set atomlist [split "$donor $acceptor"]
40   set name1 [$myseldonors get name]
41   set res1 [$myseldonors get resid]
42   set index1 [$myseldonors get index]
43   set name2 [$myselacceptors get name]
44   set res2 [$myselacceptors get resid]
45   set index2 [$myselacceptors get index]
46   set dis [measure bond $atomlist]
47   puts $chan "$molname\t$res1\t$name1\t$index1\t$res2\t$name2\t$index2\t$dis"
48   set count [expr {$count + 1}]
49   set sum [expr {$sum + $dis}]
50   }
51   set ave [expr {$sum/$count}]
52   puts $chan2 "$molname\t$count\t$ave"
53
54 close $chan
55 close $chan2
56 $
```

3.4.2. Código en Tcl Tk para obtención de contactos de van der Waals

```
1 foreach pdb [lsort [glob e_*.pdb]] {
2 mol new $pdb
3 set chan [open ${pdb}CCoutput.dat w]
4 set protC [atomselect top "(element C)"]
5 set mylist [measure contacts 4 $protC]
6
7 set mylist1 [lindex $mylist 0]
8 set mylist2 [lindex $mylist 1]
9
10 set n 0
```

```
11
12 foreach i $mylist1  j $mylist2 {
13 #puts "$i\t$j"
14 set mydist [measure bond "$i $j"]
15 set mysel0 [atomselect top "index $i"]
16 set mysel1 [atomselect top "index $j"]
17
18
19 set myresid0 [$mysel0 get resid]
20 set myname0 [$mysel0 get name]
21 set myresname0 [$mysel0 get resname]
22
23 set myresid1 [$mysel1 get resid]
24 set myname1 [$mysel1 get name]
25 set myresname1 [$mysel1 get resname]
26
27 set n [expr $n+1]
28
29 puts "$n\t$i\t$j\t$myresname0\t$myresid0\t$myname0\t$myresname1\t$myresid1\t$
    myname1\t$mydist"
30 puts $chan "$n\t$i\t$j\t$myresname0\t$myresid0\t$myname0\t$myresname1\t$myresid1
    \t$myname1\t$mydist"
31 }
32 mol delete all
33 close $chan
34 }
35 exit 0
```

3.4.3. Superposición de representaciones de Richardson

La representación de Richardson, o diagrama de cintas, es la forma gráfica de ilustrar la estructura tridimensional de las proteínas más utilizada desde que fuera desarrollada por Jane Richardson en 1981 [58]. Para la 'superposición' de representaciones de Richardson que permite la apreciación de la compactación hacia el CM de las estructuras consideradas, se procede de la siguiente manera: se superpone la representación de Richardson de la estructura de interés sobre la representación de una estructura de referencia, y se utiliza una escala de gradiente de colores azul—blanco— rojo, correspondiendo el azul a un alejamiento relativo hacia el exterior de los átomos con respecto a la posición de los átomos equivalentes en la estructura de referencia obtenida de la clase de cadenas; el blanco indica la conservación de la posición, y el rojo el

desplazamiento de los átomos en dirección al CM de la proteína. De esta forma, un diagrama o una región de diagrama donde prevalece el rojo indica que la estructura de interés se ha contraído hacia el CM, una región blanca señala que las distancias de los átomos al CM en la estructura de interés son similares a las distancias en la referencia y una región azul indica que los átomos de la estructura considerada se alejan hacia la periferia en la relación con la estructura de referencia.

Resultados

Capítulo 4

Resultados

4.1. La relación entre el R_G y la longitud de la secuencia proteica.

Se inspeccionó una subpoblación del PDB [6]. Todas las proteínas con 20 o más representaciones cristalográficas fueron incluidas en el espacio muestral. La elección de un punto de corte en el número exigido de réplicas obedeció a un criterio de robustez del análisis estadístico.

Como se describe en la sección Materiales y Métodos, a partir de un set original que incluía todas las estructuras que cumplían con el criterio reseñado y tras de una serie de pasos de depuración, fueron analizadas un total de 19,393 cadenas de proteínas, pertenecientes a 631 clases de cadenas provenientes de 7,114 entradas distintas del PDB. De dichas estructuras se obtuvo el R_G de la cadena principal y la distancia de cada uno de sus átomos pesados al CM definido, que coincide con el centro geométrico.

El R_G es un parámetro físico que guarda cierta relación con la longitud de la secuencia. La teoría clásica formulada por Flory a mediados del siglo XX [19] establece una ley potencial que relaciona ambas variables. Esta ley deducida en la física de polímeros habitualmente se expresa como:

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

$R_G = R_0 N^u$

En la ecuación N es el número de residuos, R_0 es un coeficiente que depende de detalles químicos de la molécula y la geometría de los enlaces, y u es un exponente que, de acuerdo al trabajo original de Flory, se aproxima a $0,6$ en polímero extendidos de cadena larga, donde predominan la interacciónes entre el polímero y el solvente por sobre las intrapoliméricas. Sin embargo, muchos trabajos posteriores han demostrado que, en estructuras nativas, la relación de Flory entre longitud de secuencia y R_0 se cumple, pero para valores del exponente u mucho más pequeños, de entre 0.3 y 0.4, según se considere proteínas de un solo dominio, monoméricas, oligoméricas, o el cálculo se realice basado en la unidad asimétrica de la estructura cristalina.

En nuestra investigación, calculamos el R_G de decenas de miles de proteínas y observamos que se cumple la relación anticipada por Flory, lo que se verifica en el gráfico de doble logarítmo (figura 4.1).

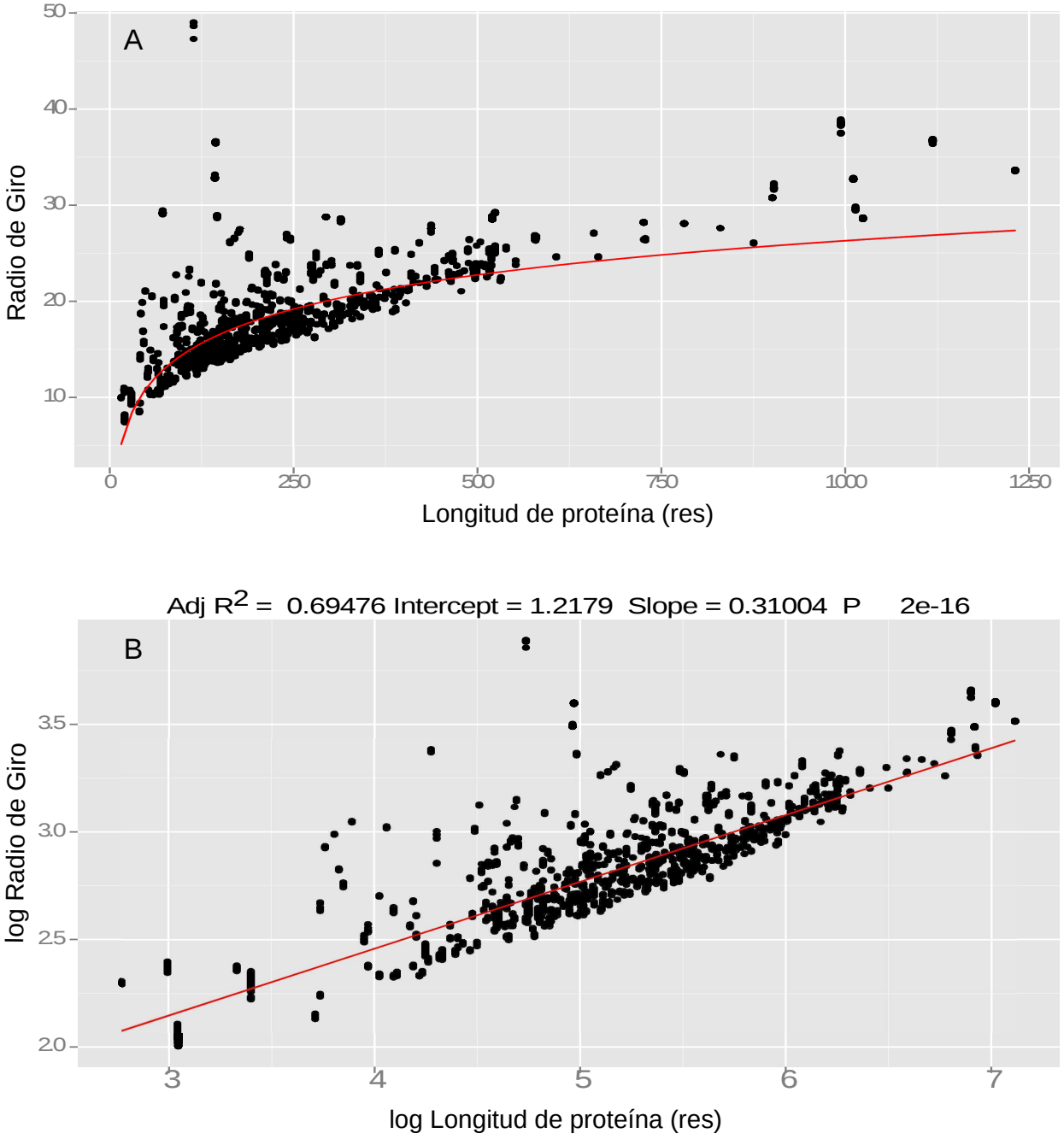


Figura 4.1: R_G **en función de la longitud de secuencia. A**: consistente con la teoría clásica de Flory, nuestro análisis de datos confirma una relación de tipo ley potencial entre el R_G y longitud de la proteína dada por $R_G = N^u$. **B**: representación logarítmica (log de R_G vs. log de longitud) sobre el conjunto total de estrucuturas analizadas. El ajuste lineal tiene pendiente $u = 0.31$, congruente con lo esperado de acuerdo a la ley potencial que vincula el R_G del esqueleto peptídico en estructuras cristalinas y la longitud de la secuencia

4.2. Contracción de las estructuras de proteínas inducida por temperatura

A fin de evaluar la influencia de la temperatura de recolección de los datos de difracción de rayos X (TCD) sobre el grado de compactación de las cadenas de proteína, se compararon estructuras de diferentes clases de cadenas (cada clase corresponde a una secuencia). A tal efecto, un valor Z fue calculado para toda las estructuras. Z es el valor normalizado de R_G que representa, en unidades de desviación estándar, la diferencia entre el R_G de una estructura particular y el valor promedio de R_G en la clase de cadenas a la que dicha estructura pertenece. Para excluir la posibilidad de que la variación de R_G sea una consecuencia de artefactos de refinamiento en la solución de la estructuras cristalográfica, se determinaron las longitudes de los enlaces peptídicos en la cadena principal, corroborándose la consistencia con los valores de distancia de enlace reportados por Linus Pauling y Robert Corey en 1951 [51]. En la tabla 4.1 se detallan la cantidad de estructuras para cada rango de temperatura y el promedio del valor de R_G estandarizado como $\bar{Z}$ para cada intervalo de temperatura.

TCD	$\leq$ 160 K	160K - 220K	$\geq$ 240 K	n.d	Total
N	14509	274	2318	2292	19393
$\bar{Z}$	-0.08	-0.01	0.37	0.12	0.00

Tabla 4.1: **Número de estructuras por rango de temperaturas y valores Z.** TCD es la temperatura de recolección de datos de difracción. Los datos de estructuras cuya temperatura no fue reportada en el PDB se encuentran en la columna 'no disponible' (n.d). N es el número de cadenas en cada rango de temperatura. $\bar{Z}$ son los valores normalizados de los promedios de R_G calculados como se describe en Materiales y Métodos. La diferencia en $\bar{Z}$ entre estructuras a ultrabaja temperatura ($\leq$ 160 K) y estructuras temperaturas moderadas ($\geq$ 240 K) es estadísticamente significativa.

Los valores de $\bar{Z}$ según el rango de TCD demuestran que las estructuras cuyos datos de difracción fueron recolectados a ultra-baja temperatura (temperaturas menores a 160 K; típi-

camente a 100 K), a las que en adelante identificaremos como 'estructuras a temperatura ultrabaja', poseen valores de R_G significativamente más pequeños que las estructuras cristalinas cuya solución se obtuvo a partir de datos de difracción recolectados a temperatura moderada (240 – 310 K), las que en adelante denominaremos simplemente 'estructuras a temperatura moderada'.

La diferencia entre los máximos de las curvas de densidad de distribución del R_G normalizado (como valor Z) de las 'estructuras a temperatura ultrabaja' y las 'estructuras a temperatura moderada' es significativa. Por lo tanto, surge evidente que el grado de compactación de las estructuras cristalográficas difiere para la misma proteína, según las condiciones de temperatura en la que se hizo la recolección de datos. No obstante, la respuesta de R_G a las variaciones de la TCD no es lineal ni es continua. En efecto, en nuestro trabajo pudimos identificar tres regiones de R_G claramente diferenciadas según la TCD. Por un lado, a ultrabaja temperatura las estructuras presentan un R_G constitutivamente más chico que a temperaturas moderadas (que incluye a las estructuras resueltas a temperatura ambiente). Pero, sugestivamente, la curva de distribución de la densidad de los R_G, tipificados como valor Z del conjunto de las cadenas que fueron resueltas con datos de difracción obtenidos a temperaturas intermedias, que son aquellas que van entre los 160 y 240 K, no expone un perfil con un máximo definido propio, sino una distribución bimodal, con máximos relativos que coinciden en su posición en el eje de los Z con los máximos respectivos de las densidades de distribución de los Radios de Giro normalizados de las 'estructuras a ultrabaja temperatura' y las 'estructuras a moderada temperatura',(figura 4.2). En la Discusión proponemos una explicación para este comportamiento que se observa en el rango intermedio de temperaturas, y que es compatible con una transición de fase con coexistencia de estados.

Asímismo, se analizó la densidad de distribución de R_G normalizado de las estructuras resueltas a temperaturas intermedias (161–220 K), evidenciando un perfil bimodal de la distribución, con sendos máximos coincidentes con los máximos de las estructuras a ultrabaja y moderada temperatura, respectivamente (figura 4.2).

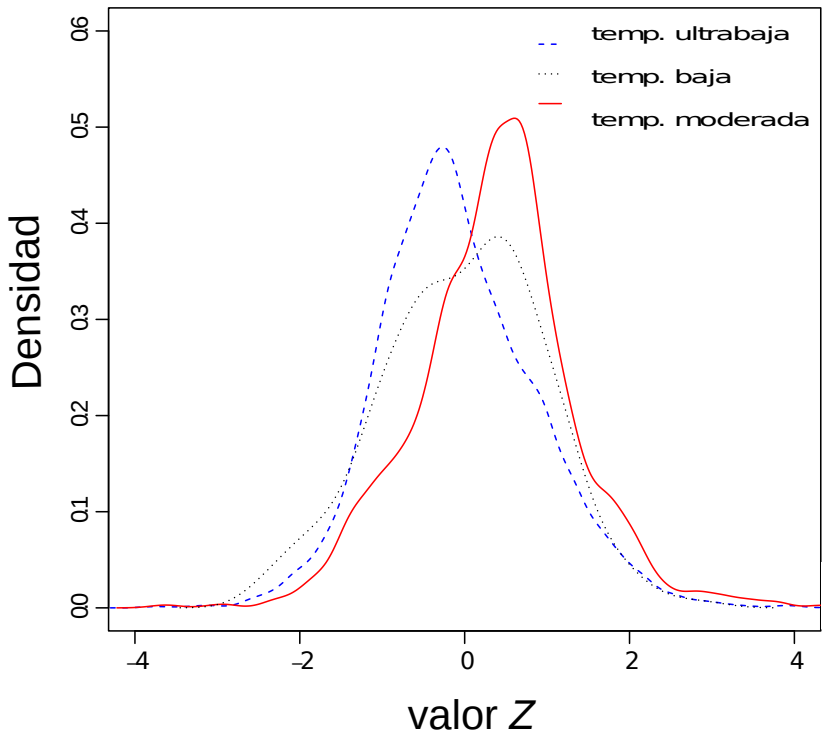


Figura 4.2: **Densidad estimada de la distribución de los valores de Z para tres diferentes rangos de temperaturas considerados.** El Valor Z mide en unidades de desviación estándar la distancia del R_G de una cadena de proteína al valor medio de los R_G de la clase de cadenas a la que pertenece. La diferencia entre los máximos de las curvas que representan las densidades de distribución de las ‘ estructuras a temperatura ultrabaja’ y las ‘estructuras a temperatura moderada’ es de 0.44 unidades de desviación estándar Z. Esta diferencia es estadísticamente significativa de acuerdo a la prueba de suma de rangos de Wilcoxon Mann Whitney ($P \leq 2.2$ x 10^{-16}), ANOVA ($F = 205.6$; $P \leq 2$ x 10^{-16}) y Tukey post HOC HSD test ($P \leq 0.05$).

4.3. Estados ultracompactos de proteínas nativas

Los resultados presentados en la sección anterior muestran que R_G es un parámetro adecuado para la detección de efectos sutiles sobre la compacidad de las proteínas. En consecuencia, se examinaron los casos que presentaban una mayor desviación a izquierda en la cola de la distribución que corresponde a los R_G normalizados de las estructuras resueltas a temperatura ultrabaja, en busca de casos con una contracción mayor a la esperada por el efecto de la temperatura.

En dicha cola de la distribución, hallamos una serie de estructuras con $Z \leq -3$. Una desviación tan extrema debe, de por sí, corresponderse con estados de la proteína de un grado de compactación muy acentuado. Sin embargo, por la forma de normalización de R_G que se representa con el valor Z, es probable que aún así esté subestimada. Esta subestimación se deduce de la forma de cálculo del valor Z de tipificación. Para obtener el valor Z se utiliza una media y una desviación estándar de R_G por clase de cadenas, sin discriminar las temperaturas de colección de datos de difracción e incorporando en el cálculo datos atípicos (*outliers*) que podrían ser descartados por criterios estadísticos de tipo Chauvenet. Un modo más realista de obtener los valores de normalización fue utilizado para recalcular Z a partir de las medias y desviaciones correspondientes a la curva de densidad de distribución de los R_G de la campana central en cada rango de temperatura.

Utilizando los valores de Z recalculados con este criterio para todas las clases de cadenas, se identificaron 23 casos con $Z \leq -4$ correspondientes a ocho clases distintas.

Las cadenas ultracompactas listadas en la tabla tienen valores de R_G mucho menores que las estructuras resueltas a ultrabaja temperatura en sus clases de cadenas respectivas. De ahí que la

Clase	**Cadena**	$T(K)$	Z
β2M humana 99	3VFN B	100	−4.3
β2M humana 99	1T1Z B	100	−4.2
β2M humana 100	3UTT B	100	−5.5
β2M humana 100	3UTT G	100	−4.3
β2M raton 100	1K8D B	100	−5.0
Cyclina A2	2WMA B	100	−8.2
Cyclina A2	2WMA A	100	−5.6
Tripsina	1QL7 A	287	−9.4
Tripsina	1AQ7 A	93	−5.0
Tripsina	2A7H A	100	−6.1
Tripsina	2G81 E	100	−7.6
Tripsina	3GY2 A	100	−12.8
Tripsina	3GY3 A	100	−12.5
Tripsina	3GY5 A	100	−13.1
Tripsina	3GY6 A	100	−12.4
Tripsina	3GY8 A	100	−12.8
Tripsina	3RU4 T	100	−15.1
Quimiotripsina A	3RU4 D	100	−9.6
Lisozima	3IJV A	100	−10.7
Fab pesada	2DWD A	100	−4.9
Fab liviana	2DWD B	100	−5.5
FC VII pesada	2FLR H	100	−6.4
HLA II DRα	2Q6W A	98	−14.8

Tabla 4.2: **Estructuras ultracompactas.** Los valores de Z fueron recalculados utilizando la media y la desviación estándar de la curva de distribución según rango de temperatura. Aunque formalmente β2M humana de 99 aminoácidos y β2M humana de 100 aminoácidos pertenecen a dos clases distintas, la única diferencia entre ellas es que la β2M humana de 100 aminoácidos incluye el residuo de metionina inicial. La tripsina 1QL7 pertenece al rango de las estructuras resueltas a temperaturas moderadas, pero fue incluida porque exhibe un grado alto de compactación.

contracción térmicamente inducida -descrita anteriormente- no alcanza para justificar la ultracompactación. Las posibles explicaciones triviales y artefactuales fueron descartadas mediante un riguroso control de calidad de las estructuras incluidas en el análisis (ver Materiales y Métodos). Asimismo, las variaciones no son atribuibles a diferencias en posiciones rotaméricas de las cadenas laterales, toda vez que en el procedimiento de cálculo de R_G se consideró solamente

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

las coordenadas de los átomos de la cadena principal en estructuras de secuencia indéntica (ver Materiales y Métodos).

Un examen más detenido de las estructuras listadas en la tabla 4.2 confirma que los cambios conformacionales involucrados en la disminución del R_G de estas cadenas tienen características distintas a los antecedentes referidos en la literatura citada en la sección de introducción. Tal como se describe en Materiales y Métodos, para asegurarnos la discriminación entre variaciones de R_G secundarias a una eventual reorganización de subdominios mediante movimientos de cuerpo rígido o desplazamientos de tipo bisagra, sólo admitimos dentro de la categoría de 'cadenas ultracompactas' a estructuras cuya marcada desviación a izquierda del R_G fuera acompañada con un movimiento global de los átomos en dirección al CM o centro geométrico de la estructura ($\geq$ 80 % de los átomos). Este modo de compactación de carácter isotrópico no modifica la forma de la proteína sino el volumen y la proteína ultracompacta preserva la posición relativa de los dominios. En otro orden, no se detecta un acortamiento de las distancias de los enlaces peptídicos carbono — nitrógeno, más allá del entorno de la longitud típica de 1.32 Å determinadas por Linus Pauling [51], lo que sería un fuerte indicio de un artefacto de refinamiento. En consecuencia se observa una genuina contracción global de estas proteínas sin perjuicio de la estructura covalente.

A continuación se analizó el comportamiento estadístico de las interacciones no covalentes. En promedio los enlaces de hidrógenos son más cortos en la estructuras resueltas a ultrabaja temperatura ($\leq$ 160 K) que en las estructuras resueltas a estructura moderada ($\geq$ 240 K), alcanzando una media de 3.041 Å contra 3.051 Å respectivamente, con significación estadística evaluada por test de student (P $\leq 3.2 \times 10^{-18}$). La misma comparación efectuada entre las estructuras resueltas a temperatura ultrabaja y las estructuras ultracompactas mostró una

disminución más notable de la longitud los puentes de hidrógeno: 3.041 Å contra 2.971 Å, respectivamente ($P \leq 3.4 \times 10^{-8}$).

Por su parte, en promedio los contactos de van der Waals entre átomos de carbono mantienen una distancia similar en las estructuras resueltas a ultrabaja temperatura y en las estructuras resueltas a temperatura moderada. Sin embargo, cuando se comparan las estructuras de temperatura ultrabaja con las estructuras ultracompactas, que en la práctica son una subpoblación de las primeras, se observa un acortamiento en la distancia media de las interacciones de van der Waals: 3.746 Å en ultrabaja temperatura, mientras que 3.737 Å en ultracompactas, con P= 0.0002).

Para poder comparar el número de interacciones no covalentes de las estructuras según la TCD, obtuvimos un valor Z de los datos, utilizando el mismo procedimiento de normalización aplicado a los valores de R_G . El análisis develó que el número de puentes de hidrógeno en las estructuras cristalinas no es afectado por la temperatura ni por el grado de compactación estimada por el R_G. Por el contrario, la cantidad de contactos de van der Vaals entre átomos de carbono aumenta de forma significativa cuando la recolección de datos de difracción se realiza a temperatura por debajo de 160 K, y todavía más en el caso de estructuras que muestran ultracompactación, según R_G.

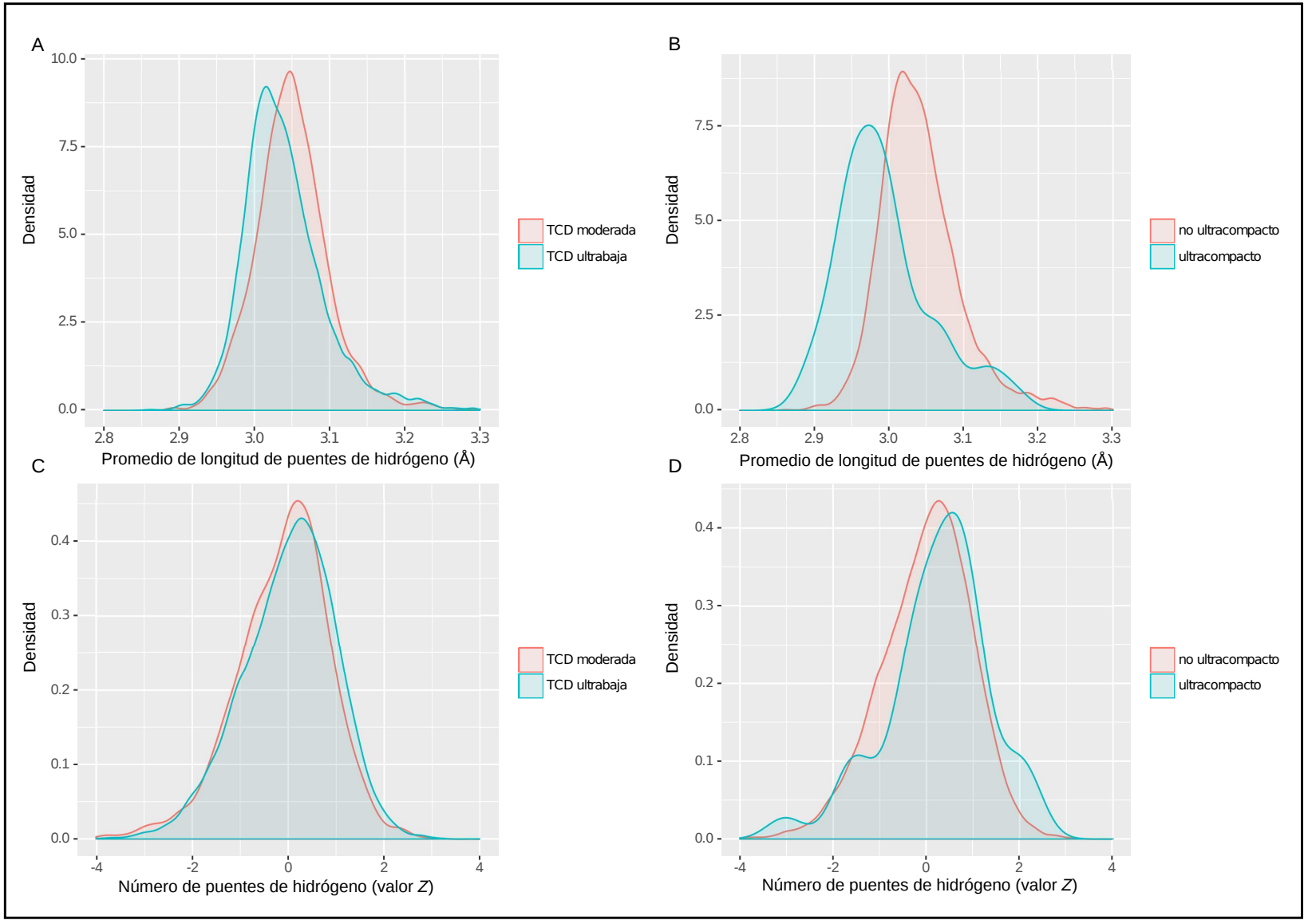


Figura 4.3: **Puentes de Hidrógeno según TCD y grado de compactación.** El número de puentes de hidrógeno no covaría con R_G ni es sensible a TCD, pero sus longitudes medias sí los son.

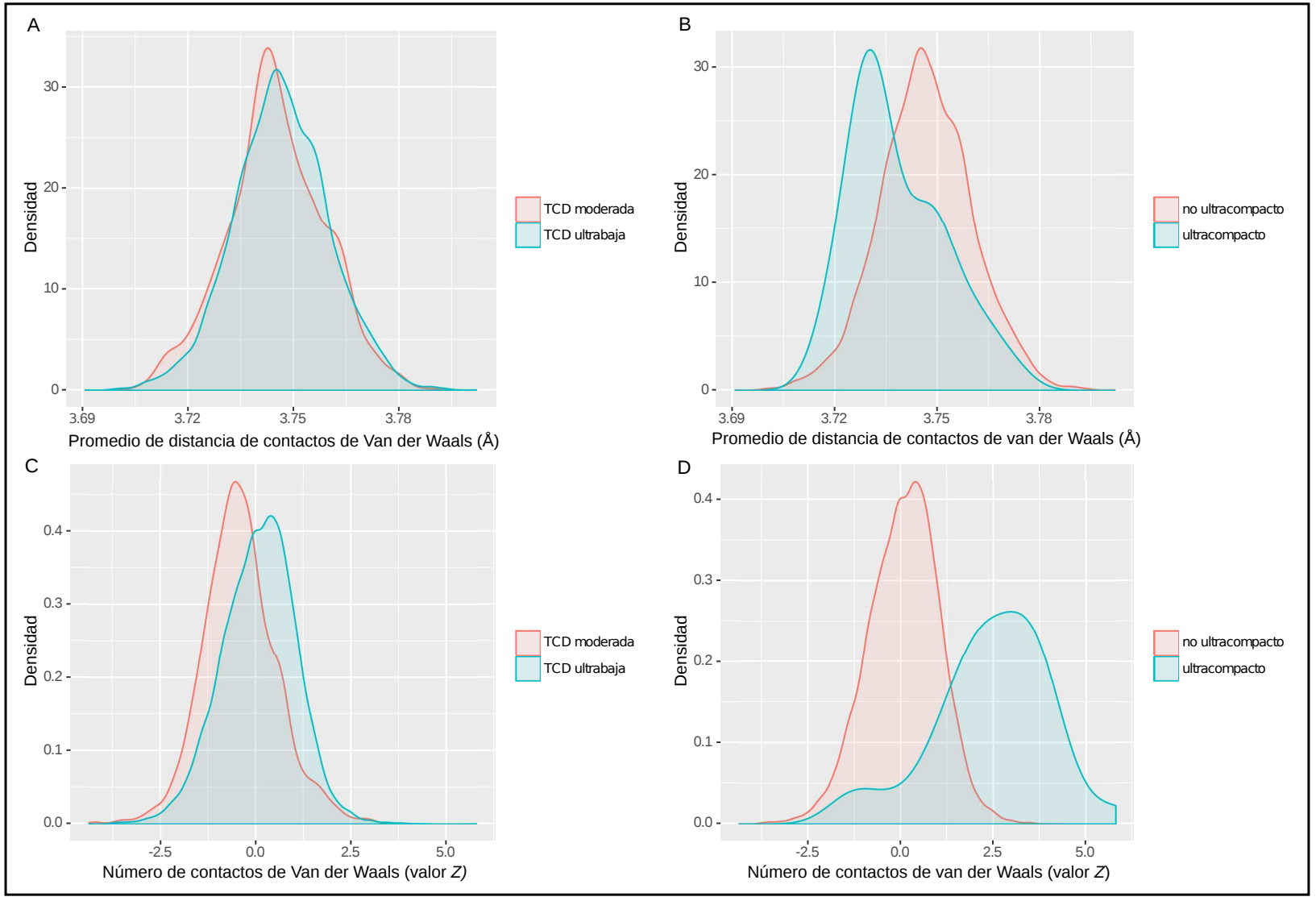


Figura 4.4: **Contactos de van der Waals según TCD y grado de compactación.** El número de contactos de van der Waals aumenta con el grado de compactación y es sensible a TCD. Conforme disminuye R_G, las interacciones de tipo de van der Waals entre átomos de carbono se estrechan.

4.4. Tripsina

En la clase de cadenas de la tripsina bovina (UniProtKB P00760) se incluyeron 261 estructuras cristalográficas, entre las cuales se identificaron una decena de ultracompactas (tabla 4.2). En la figura 4.5 (A) se aprecia que los valores de R_G de la población se distribuyen en una región muy estrecha en torno a $R_G = 15.83$ Å, para más del 95 % de las tripsinas depositadas en el PDB.

Tras un primer examen de la curva de densidad, se observa que ocho estructuras exhiben valores de R_G que se desvían hacia la izquierda del promedio de la distribución más allá de la línea de corte de 3 σ que exigimos para catalogarlas como ultracompactas.

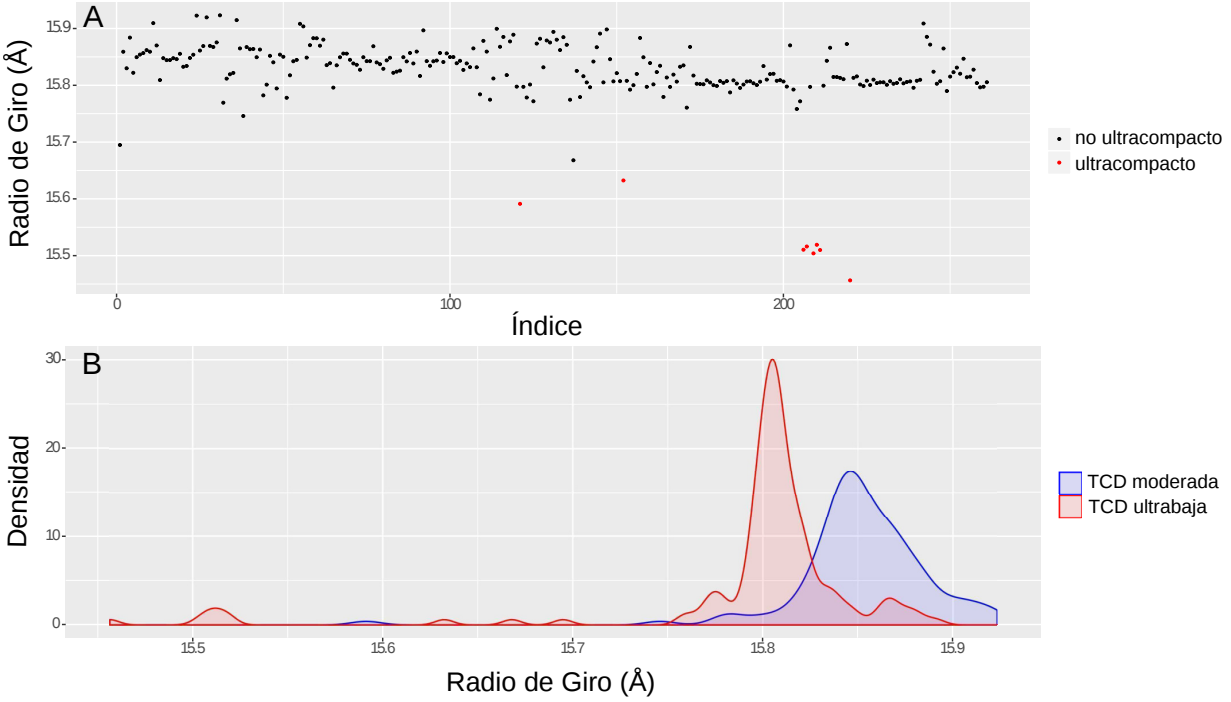


Figura 4.5: **Distribución de R_G de tripsinas y densidad de la distribución de R_G según TCD.** **A:** más del 95 % de las tripsinas depositadas en el PDB se distribuyen en torno a un $R_G = 15.83\mathring{A}$. En rojo se destacan la tripsinas cuyo R_G está desplazado al menos 3 σ hacia la izquierda, considerando el valor medio y la desviación estándar de la totalidad de las cadenas de la clase. **B:** la densidad de R_G según la temperatura de colección de datos muestra que las tripsinas ultrafrías tienen R_G menor y la mayoría de las ultracompactas son ultrafrías

.

Dos hechos sugieren que la compactación de estas estructuras tiene una significación estadística más importante de la que se estima al utilizar el valor de la desviación estándar σ y la media μ de la variable R_G que se obtiene sobre el total de las cadenas en la clase de las tripsinas.

En primer lugar, salvo una de las cadenas, todas las tripsinas ultracompactas son 'estructuras a temperatura ultrafría'. Dichas estructuras conforman una subpoblación estadística dentro de la clase para la variable R_G, puesto que en las tripsinas se reproduce el comportamiento general de compactación inducida por la TCD (figura 4.5,(B)). En segundo término, es evidente que la línea de corte de 3σ de alejamiento implica que las cadenas ultracompactas son datos atípicos tan extremos (*outliers*) que pueden considerarse una población distinta y sus valores de R_G no se deben incluir en los cálculos de los parámetros estadísticos (μ y σ) de la distribución mayoritaria, porque producen resultados engañosos. En consecuencia, el alejamiento de los datos medido desde la μ y en valores de σ de la población ultrafría típica (sin *outliers*) refleja de forma más apropiada el fenómeno que investigamos (ver figura 4.6).

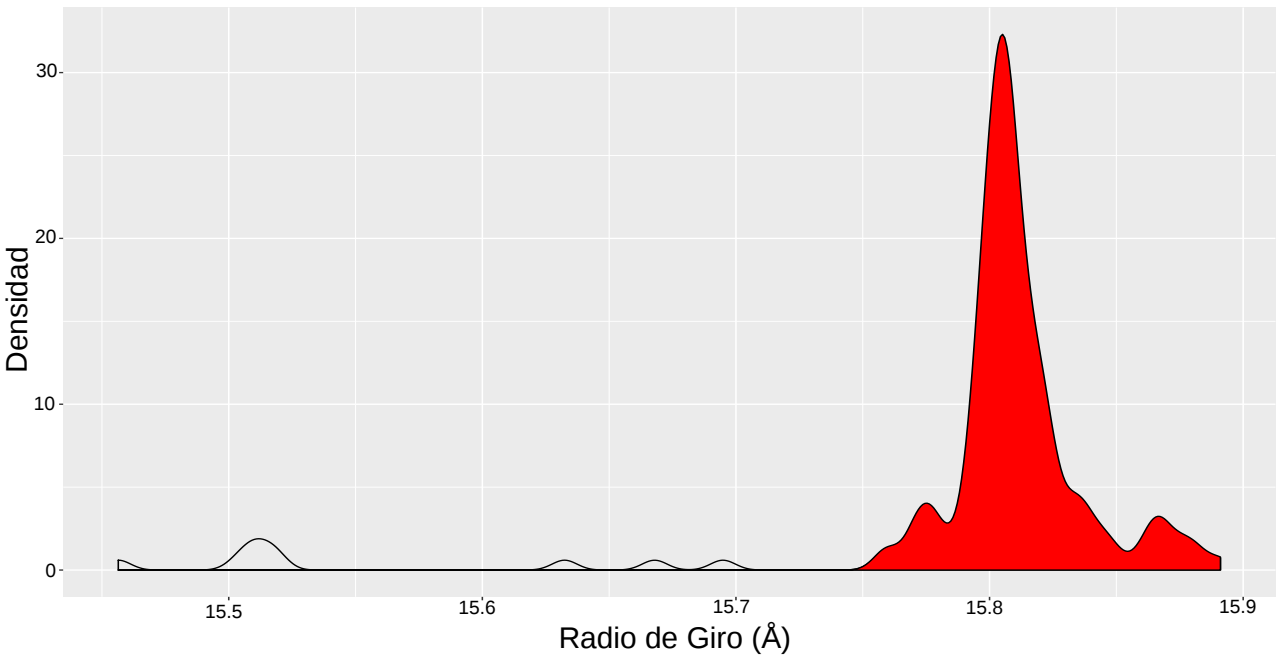


Figura 4.6: **Densidad de la distribución de R_G en tripsinas ultrafrías.** La densidad de distribución corresponde a las tripsina con TCD ultrabaja. La región de la densidad destacada en color rojo (campana central) fue la utilizada para obtener la media y la desviación estándar de la población con la que se calculó el Valor Z definitivo.

Con este nuevo procedimiento, hasta diez estructuras cumplieron un criterio de ultracompactación todavía más exigente (línea de corte en -4σ) en virtud de las ubicaciones relativas de sus radios de giro con respecto al valor medio de la distribución (ver tabla 4.4).

PDB ID	Átomos	R_G	C—N	TCD	Valor Z 1	Rango TCD	Compactación	Valor Z 2
1QL7 A	892	15.59	1.32	287.00	-3.75	TCD moderada	ultracompacto	-6.97
1AQ7 A	892	15.69	1.33	93.00	-2.09	TCD ultrabaja	no ultracompacto	-4.96
2A7H A	892	15.67	1.33	100.00	-2.52	TCD ultrabaja	no ultracompacto	-6.10
2G81 E	892	15.63	1.32	100.00	-3.09	TCD ultrabaja	ultracompacto	-7.61
3GY2 A	892	15.51	1.33	100.00	-5.04	TCD ultrabaja	ultracompacto	-12.79
3GY3 A	892	15.52	1.32	100.00	-4.95	TCD ultrabaja	ultracompacto	-12.55
3GY5 A	892	15.50	1.32	100.00	-5.14	TCD ultrabaja	ultracompacto	-13.06
3GY6 A	892	15.52	1.32	100.00	-4.90	TCD ultrabaja	ultracompacto	-12.42
3GY8 A	892	15.51	1.33	100.00	-5.05	TCD ultrabaja	ultracompacto	-12.81
3RU4 T	892	15.46	1.32	100.00	-5.91	TCD ultrabaja	ultracompacto	-15.08

Tabla 4.3: **Estructuras ultracompactas de tripsina.** La columna Átomos refiere al número de átomos distintos del hidrógeno en la cadena principal. La columna C—N corresponde a la longitud promedio de los enlaces peptídicos entre carbono y nitrógeno en Å. La existencia de dos valores Z (Z 1 y Z 2) se explica en el texto en detalle. El rango de TCD indica que TCD menores de 160 K son ultrabajas y mayor a 240 K son moderadas

El mismo procedimiento se practicó en todas las clases donde se hallaron estructuras ultracompactas de acuerdo al alejamiento de los valores de R_G, sin discriminar, a priori, las poblaciones según la TCD ni descartar los *outliers* para obtener los valores de μ y σ de la variable. (ver Figura 4.7)

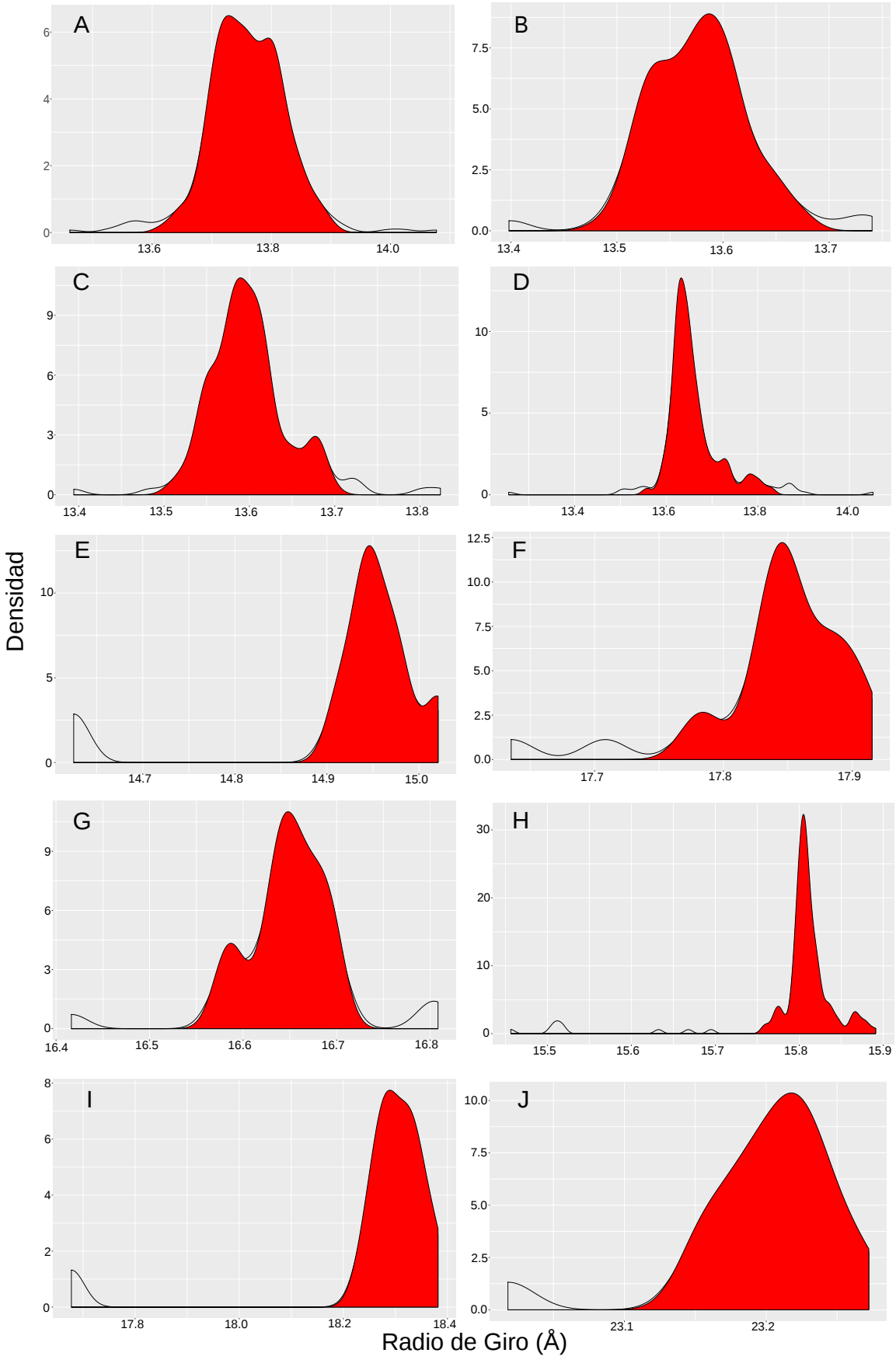


Figura 4.7: **Densidad de la distribución de R_G en clases de cadenas ultrafrías. A**: Beta 2 microglobulina humana (100 aa). **B**: Beta 2 microglobulina humana (99 aa). **C**: Beta 2 microglobulina murina. **D**: Lisozima. **E**: Quimotripsina. **F**: Ciclina A2. **G**: Factor de coagulación VII **H**: Tripsina. **I**: HLA II. **J**: Cadena liviana de Fab.

El valor atípico de R_G se debe a una compactación genuina y no a movimientos relativos de dominios ni a reorganizaciones de alcance local con efecto sinérgico sobre la distribución general de los átomos alrededor del CM. Para corroborarlo, comparamos la distancia de cada átomo de la cadena principal al CM ,o centro geométrico, de la estructura ultracompacta a la que pertenecen contra el promedio de las distancias al CM de los átomos correspondientes en el conjunto de las estructuras de toda la clase.

Este análisis comparativo puso de manifiesto que la ultracompactación se acompaña de una contracción isotrópica: la mayoría de los residuos se aproximan a las coordenadas del CM (ver tabla 4.4).

PDB ID	Átomos	R_G (Å)	Compactación (% R_G)	Aproximación (átomos)	Factor de aproximación (%)
1QL7 A	892	15.59	1.48	861	96.52
1AQ7 A	892	15.69	0.83	684	76.68
2A7H A	892	15.67	1.00	826	92.60
2G81 E	892	15.63	1.22	834	93.50
3GY2 A	892	15.51	1.99	884	99.10
3GY3 A	892	15.52	1.96	880	98.65
3GY5 A	892	15.50	2.03	885	99.22
3GY6 A	892	15.52	1.94	885	99.22
3GY8 A	892	15.51	2.00	880	98.65
3RU4 T	892	15.46	2.33	882	98.88

Tabla 4.4: **Aproximación al CM en tripsinas ultracompactas.**Los datos corresponden a las cadenas de tripsina con R_G desviado al menos 4σ hacia la iquierda de la distribución. La columna Átomos corresponde al número de átomos totales de la cadena principal sin hidrógenos. Compactación es el valor porcentual de acortamiento de R_G con relación a la media de la distribución. La Aproximación es la cantidad de átomos de la cadena principal de las estructuras ultracompactas que se acercan al centro de masa con respecto a la distancia media de los átomos respectivos en la clase. El Factor de aproximación es el porcentaje que representa la Aproximación de los átomos totales

El primer caso que se reseña en la tabla 4.4 corresponde a la cadena A de la entrada del PDB *1QL7* ($Z = -6.9$). Este es el único caso en la clase de las tripsinas de una estructura ultracompacta que proviene de la población resuelta a temperatura moderada (287 K). Figura 4.8

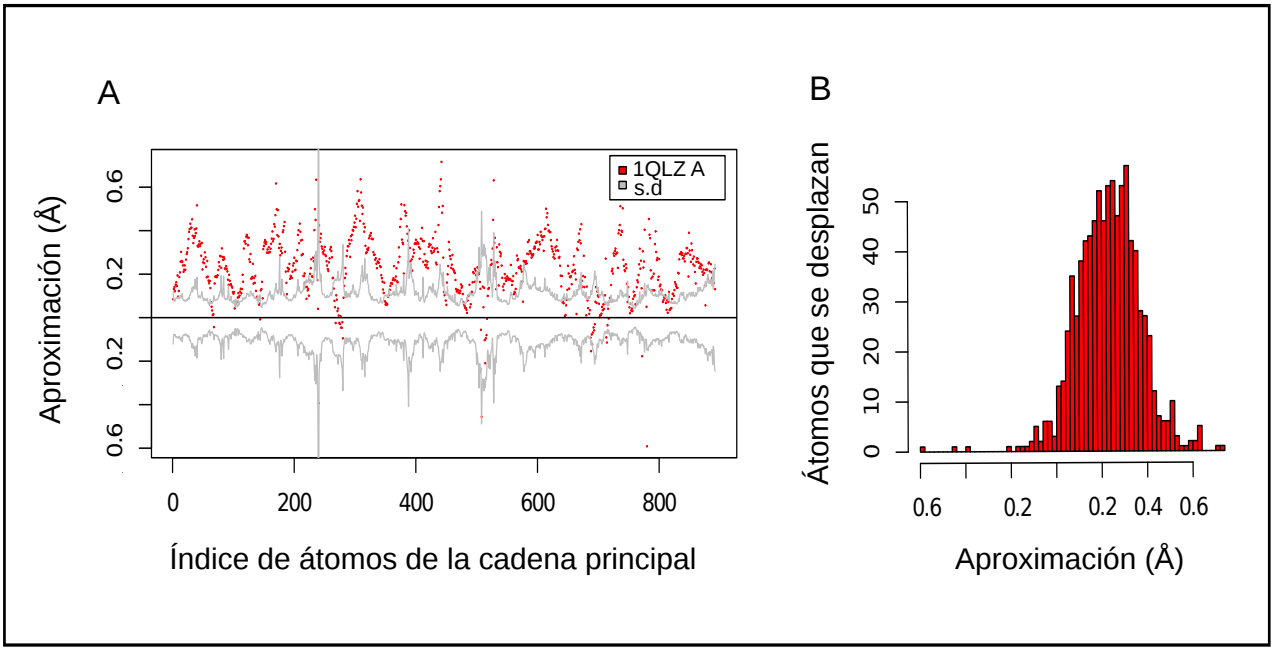


Figura 4.8: **Diferencias de distancias a CM de los átomos de la cadena principal en 1QL7.** **A**: se calcularon las distancias en Å al CM de los átomos de la cadena principal para la tripsina 1QL7. En color se representan con respecto al promedio de las distancias a CM de los átomos respectivos en la población de estructuras a temperatura moderada. Las líneas en gris grafican ± s.d en cada posición de la cadena principal. Valores positivos indican aproximación a CM; valores negativos indican alejamiento; y el valor **0** indica conservación de la distancia. **B**: histograma de diferencias de distancias al CM con respecto al valor promedio. La contracción es evidente de la distribución. Si no hubiese contracción genunina, el histograma se distribuiría en forma simétrica en torno al cero [22].

Otra forma de hacer visible la contracción de las cadenas es la representación de Richardson o diagrama de cintas [58, 59] de las estructuras ultracompactas 'superpuestas' con estructuras que no lo son, pero cuyas condiciones experimentales de resolución son similares, sobre todo en la TCD. Vale decir que, en rigor, no se realiza una superposición de una estructura sobre otra sino un representación de Richardson de la estructura de interés, coloreada de acuerdo a una escala que refleja la diferencia entre la distancia 'punto a punto' (esto es: átomo por átomo) al CM con respecto a la estructura de referencia (figura 4.9).

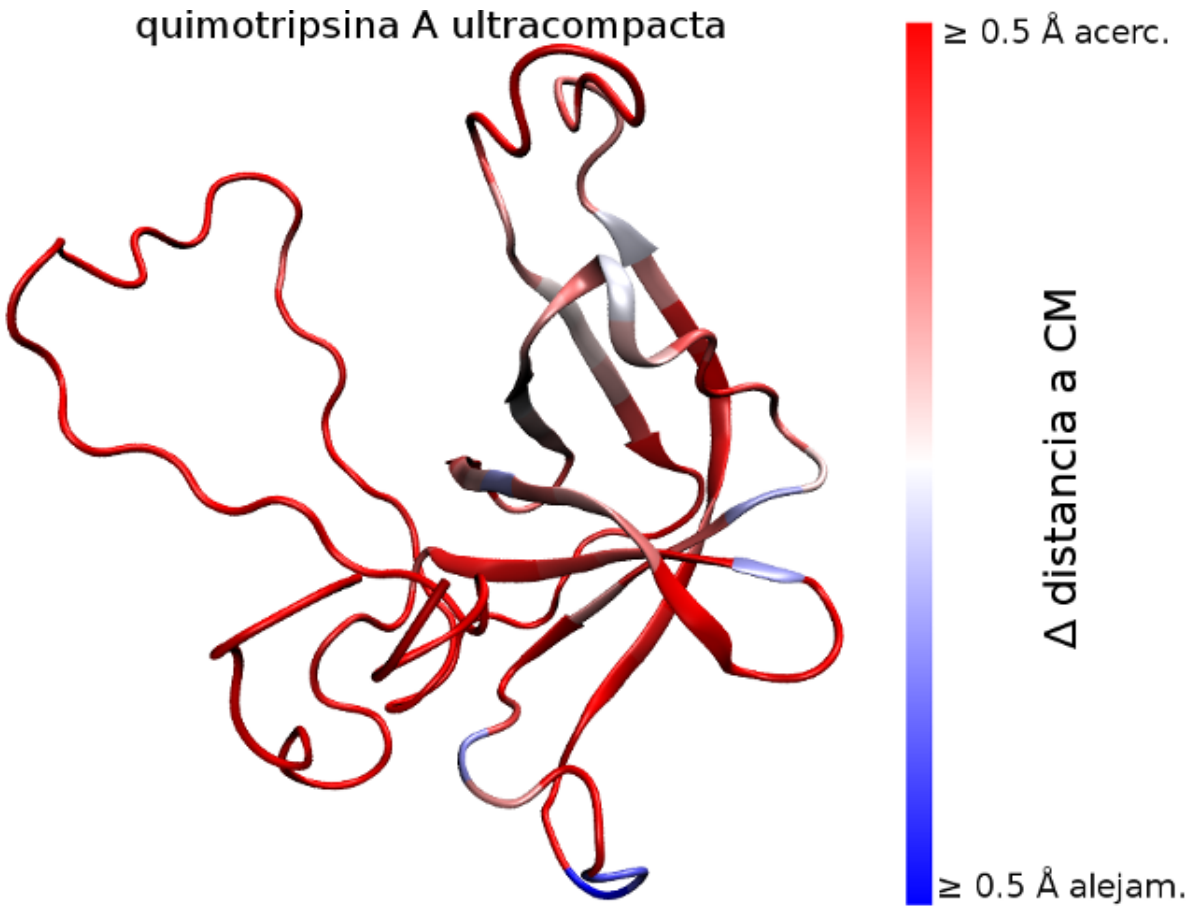


Figura 4.9: **Representación gráfica de la compactación con diagrama de Richardson.** El ejemplo muestra la escala de colores utilizada en el diagrama de cintas sobre una quimotripsina ultracompacta 'superpuesta' con una estructura de referencia

Por ejemplo, la cadena de tripsina de 1QL7 fue comparada con la cadena de tripsina 1GI6 ($Z = -0.002$) que también corresponde a una estructura de TCD moderada (298 K), pero con el valor de R_G que más se aproxima al promedio de esta subpoblación. En el panel **A** de la figura 4.10 puede observarse el resultado de la 'superposición' en una escala de colores que recorre el rojo (acercamiento relativo al CM) — blanco (conservación de distancia a CM) — azul (alejamiento relativo del CM).

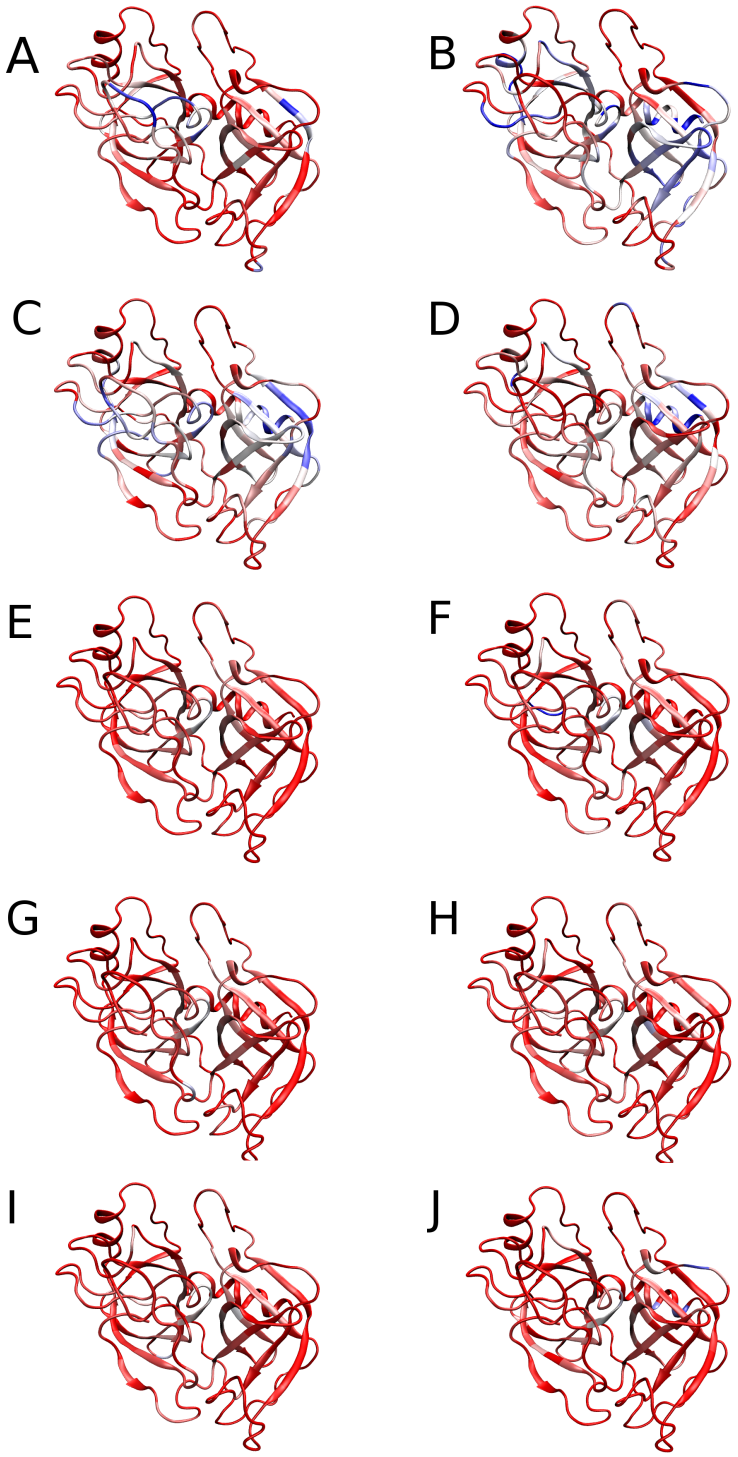


Figura 4.10: **Representaciones de Richardson de tripsinas ultracompactas 'superpuestas' con estructuras de referencia.** La diferencia de distancia al CM se representa en la escala de colores rojo — blanco — azul en el rango ≥ 0.5Å a ≤ -0.5Å. La mayoría de los átomos de cadenas ultracompactas están más próximos (rojo) al CM que en cadenas de referencia. Sólo algunos segmentos se alejan (azules) o conservan la distancia (blanco). ***Panel A.*** PDB ID: 1QL7 *vs* 1GI6; ***Panel B.*** PDB ID: 1AQ7 *vs* 1P2K; ***Panel C.*** PDB ID: 2A7H *vs* 1P2K;***Panel D.*** PDB ID: 2G81 *vs* 1P2K; ***Panel E.*** PDB ID: 3GY2 *vs* 3GY4; ***Panel F.*** PDB ID: 3GY3 *vs* 3GY4; ***Panel G.*** PDB ID: 3GY5 *vs* 3GY4; ***Panel H.*** PDB ID: 3GY6 *vs* 3GY4; ***Panel I.*** PDB ID: 3GY8 *vs* 3GY4; ***Panel J.*** PDB ID: 3RU4 *vs* 1P2K.

Es apreciable el movimiento del conjunto de los átomos de la cadena principal de la tripsina 1QL7 hacia el CM. Algunas pequeñas regiones en blanco y azul revelan la conservación de la posición (blanco) o movimientos de algunos átomos hacia el exterior (gradientes de azul), pero ello no alcanza para compensar la compactación global que se manifiesta en el valor de R_G.

En el panel **B** de la figura 4.10 se representa la superposición de la cadena de tripsina con identificación de PDB: 1AQ7 ($Z = -5.0$) y 1P2K ($Z = 0$). La estructura 1AQ7 es un complejo formado por la tripsina y el péptido inhibidor modificado aeuroginosina 98-B; mientras tanto, 1P2K describe el complejo entre la tripsina y la aprotinina, más calcio y el ión sulfato. Las dos estructuras fueron resueltas a temperatura ultrabaja y la estructura de referencia 1P2K fue seleccionada porque su valor de $Z = 0$ la ubica en la media exacta de la distribución de los R_G para su población. El desplazamiento relativo de la mayoría de los átomos de la cadena principal al CM es todavía evidente en la escala de colores, pero 1AQ7 es, entre todas las tripsinas ultracompactas, la que presenta menor factor de aproximación.

Las tripsinas 1AQ7 y 2A7H (panel **C** de la figura 4.10) conforman el grupo de tripsinas ultracompactas que no fueron detectadas en primera instancia, sino luego del análisis estadístico más fino que se explicó antes. El diagrama de cintas en las superposiciones comparativas (figura 4.10, panel **A** y **B**) y las gráficas e histogramas de aproximación a CM de la figura 4.11 demuestran la compactación global de estas estructuras. En el caso de 1AQ7, los pocos residuos que se alejan del CM lo hacen dentro del rango de una desviación estándar (*sd*), con excepción del ácido aspártico 145, cuyo alejamiento va más allá de la *sd*, como se aprecia en la posición de los cuatro atómos del backbone de ese aminoácido alrededor del índice 580, que corresponde al C_α de Asp 145.

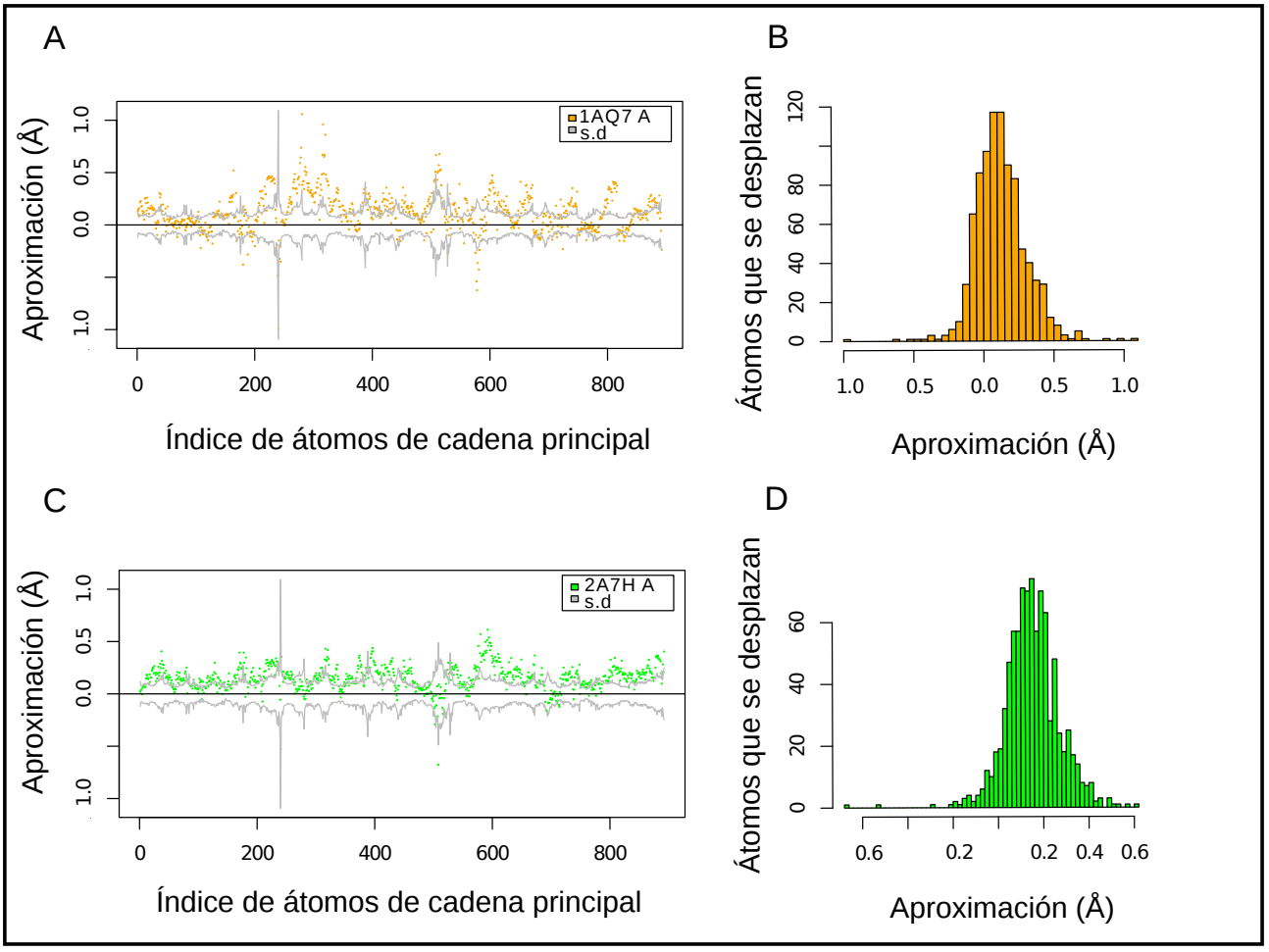


Figura 4.11: **Diferencias de distancias al CM en tripsinas ultracompactas 1AQ7 Y 2A7H.** En **A** y **C** se calcularon las distancias en Å al CM de los átomos de la cadena principal para las tripsina 1AQ7 y 2AH7, respectivamente. La mayoría de los átomos se aproximan al CM con respecto al promedio de distancia en la población de estructuras resueltas a temperatura ultrabaja. Las líneas en gris grafican ± *sd* en cada posición de la cadena principal. Valores positivos indican aproximación a CM; valores negativos indican alejamiento; y el valor 0 indica conservación de la distancia. En **B** y **D** la distribución de los histogramas hacen evidente la compactación. Si no hubiese contracción genunina en las tripsinas analizadas, los histogramas se distribuirían en forma simétrica en torno al valor central cero [22].

Las tripsinas de los identificadores PDB 2G81 y 3RU4 son cadenas ultracompactas resueltas con datos colectados a temperatura ultrafría. Se corresponden con tripsinas en complejo con un inhibidor de la familia de los inhibidores Bowman Birk de serina proteasas (BCTI), procedentes del poroto caupí (*Vigna unguiculata*). Los miembros de esta familia de inhibidores

habitualmente presentan dos sitios de inhibición, mediante los cuales interactúan con la enzima de forma no covalente. Mientras 2G81 es un complejo simple entre tripsina y BCTI, 3RU4 es un complejo triple entre tripsina — BCTI — quimiotripsina, con BCTI inhibiendo a las dos proteasas simultáneamente. Tanto en el patrón de colores que se aprecia en la figura (panel **D** y **J** de figura 4.10) como en los gráficos de aproximación (figura 4.12) se manifiesta la compactación que es significativa para las dos estructuras, pero bastante más apreciable en 3RU4, a pesar de estar inhibidas por el mismo compuesto inhibidor y pertenecer a la misma subpoblación según la TCD.

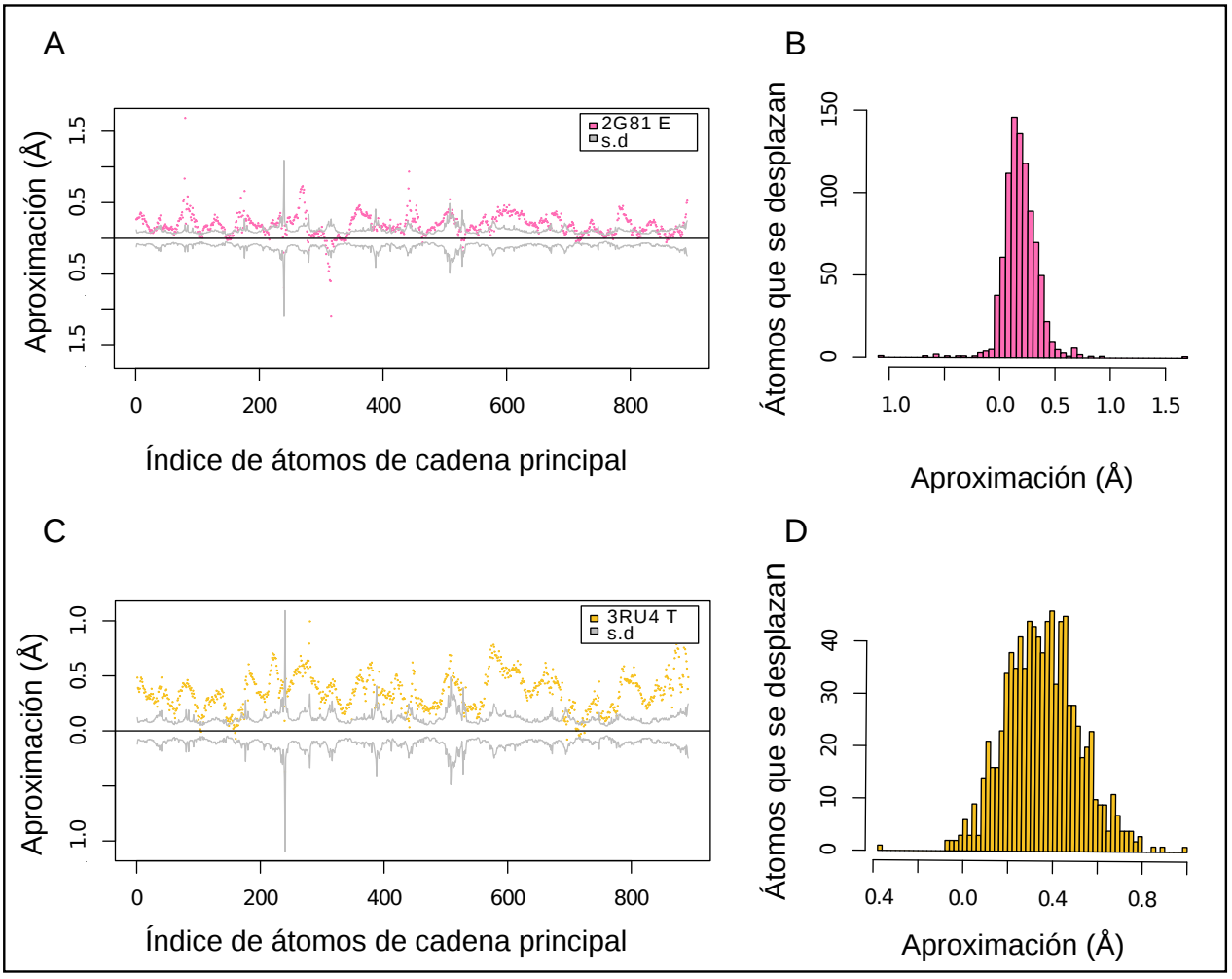


Figura 4.12: **Diferencias de distancias al CM en tripsinas ultracompactas 2G81 Y 3RU4.** **A** y **C**: aproximación a CM de los átomos de la cadena principal para las tripsinas 2G81 y 3RU4. **B** y **D**: histograma de las distancias de aproximación. El factor de aproximación es cercano a 100 en ambas estructuras, pero 3RU4 tiene un R_G mucho más pequeño, y el desplazamiento medio de sus átomos al CM es más importante.

En la figura (paneles **E** a **I** de figura 4.10) se representa la superposición de las tripsinas ultracompactas 3GY2, 3GY3, 3GY5, 3GY6 y 3GY8 con la tripsina de referencia 3GY4, y el gradiente de color en función de la aproximación al CM. El grado de compactación de todas estas estructuras es muy similar, con un marcado desplazamiento de los átomos de sus cadenas principales a los CM respectivos. Siete estructuras (3GY2 – 3GY8) fueron resueltas por el mismo grupo de investigación, en similares condiciones experimentales. Cinco de ellas

son ultracompactas, en tanto 3GY4 y 3GY7 presentan un tamaño normal.

Estas siete tripsinas están unidas a compuestos inhibidores derivados de la benzamidina, un inhibidor competitivo reversible de las serina proteasas que mimetiza la cadena lateral básica protonada de la arginina, que es un ligando natural de este tipo de enzimas. Las cadenas 3GY2, 3GY5, 3GY6 Y 3GY8 se encuentran en complejo con diminazeno, mientras que 3GY3 está unida a la pentamidina. Estos dos compuestos son diamidinas aromáticas; es decir, están formados por dos extremos benzamidina unidos por un enlazador (*linker*), y se utilizan desde hace muchos años para el tratamiento de algunas formas de tripanosomiasis **??**. El motivo de unión entre las dos amidinas aromáticas de estos compuestos es una cadena de dioxipentano en la pentamidina, y un grupo triazeno, más corto y químicamente distinto, en el diminazeno. Por su parte, 3GY4 Y 3GY7 están asociadas con benzamidina y p-amino benzamidina, respectivamente (figura 4.13).

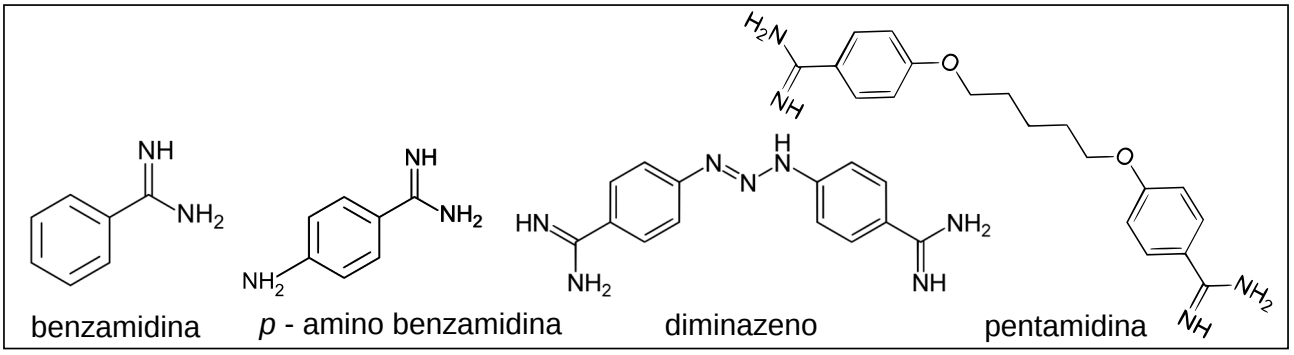


Figura 4.13: **Estructura de benzamidinas y diamidinas aromáticas.** Los compuestos derivados de la benzamidina son inhibidores competitivos reversibles de la tripsina, tradicionalmente utilizados para el tratamiento de tripanosomiasis.

La diferencia más evidente entre el grupo de las cinco cadenas ultracompactas con el grupo de las dos que no lo son, es que el inhibidor utilizado en las dos estructuras que exhiben una compactación normal es del tipo mono - benzamidina (con un solo motivo amidino aromático),

mientras que las estructuras ultracompactas están inhibidas con una bis - benzamidina (dos grupos amidino aromáticos), que apenas difieren en las características química y longitud del motivo de unión, pero no en la identidad de los grupos en sus extremos, responsables de la inhibición competitiva.

En los gráficos de aproximación al CM de este grupo de tripsinas ultracompactas no solamente se observa que se produce una contracción general en virtud del movimiento de la mayoría de los átomos hacia los centros de masa respectivos, sino que tras una inspección más fina se aprecia que la aproximación 'punto a punto' es muy similar. En otras palabras, las estructuras se ultracompactan desplazando los mismos átomos y en igual medida (ver figura 4.14). Para corrobar esta hipótesis se reprodujeron los gráficos de aproximación a CM de las cadenas ultracompactas, pero en pares con 3GY2, de forma de comparar los acercamientos, y se obtuvo la representación en la escala de colores sobre el diagramas de cintas, utilizando en este caso una estructura ultracompacta como referencia (ver figuras 4.15 y 4.16).

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

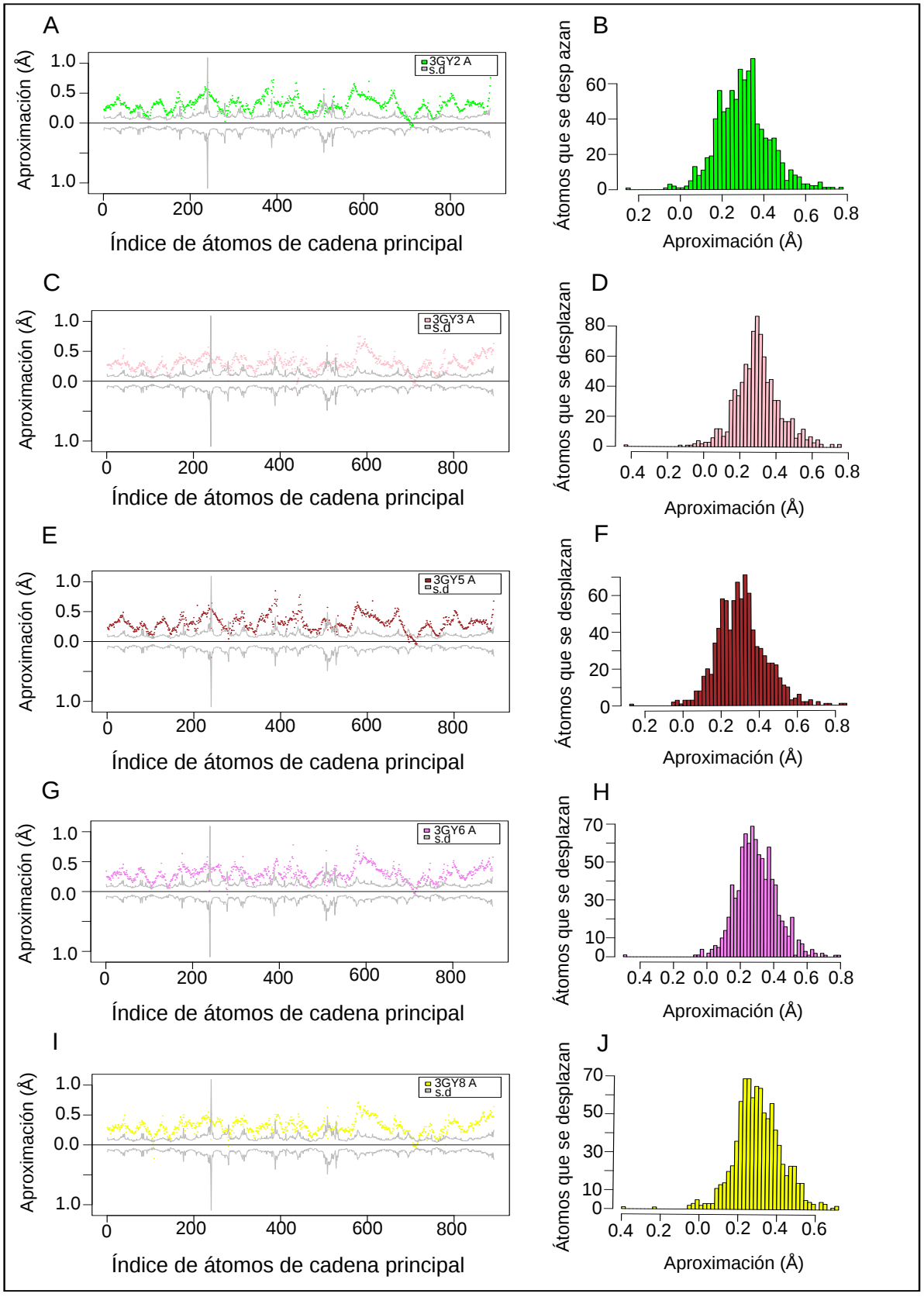


Figura 4.14: **Diferencias de distancias al CM en tripsinas ultracompactas 3GY2, 3GY3, 3GY5, 3GY6 y 3GY8.** Figuras **A**, **C**, **E**, **E** e **I**: aproximación a CM en los átomos de la cadena principal de las cadenas respectivas. Figuras **B**, **D**, **F**, **H**, y **J**: histograma de datos de aproximación. La compactación es apreciable en todos los casos. Las curvas de los datos de aproximación resultan similares, especialmente entre las tripsinas inhibidas con diminazeno.

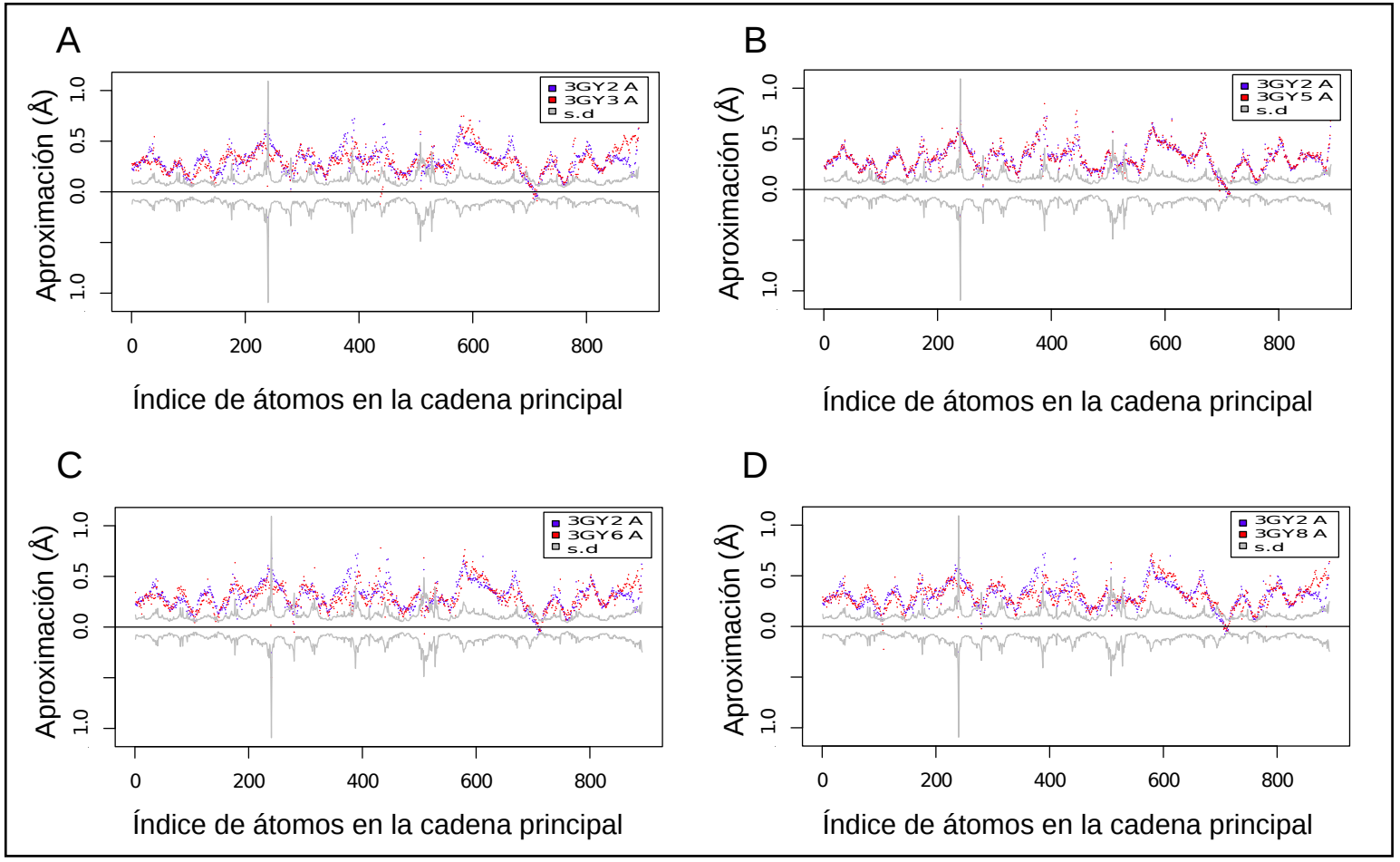


Figura 4.15: **Gráfico de diferencia de distancia al CM.** La aproximación al CM de los átomos de la cadena principal de en las cadenas de interés, relativa al promedio de distancias al CM de la clase a la que pertenecen se destaca en color. Los gráficos de aproximación superpuestos de 3GY2 con el resto de las tripsinas 3GYx ultracompactas demuestran que en todas las tripsinas de este grupo se desplazan los mismos átomos y en similar medida hacia el CM. **A**: perfil de aproximación de 3GY2 y 3GY3. **B**: perfil de aproximación de 3GY2 y 3GY5. **C**: Perfil de aproximación de 3GY2 y 3GY6. **D**: Perfil de aproximación de 3GY2 y 3GY8.

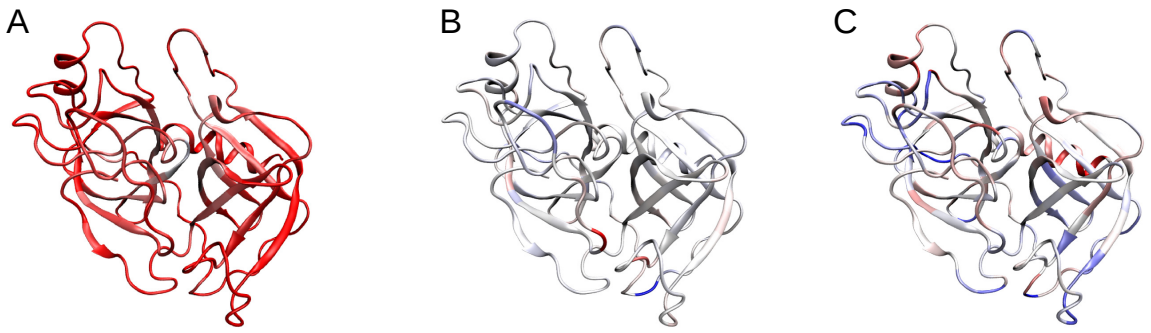


Figura 4.16: **Representaciones de Richardson de tripsinas ultracompactas superpuestas entre ellas.** La diferencia de distancia al CM se representa en la escala de colores rojo — blanco — azul en el rango ≥ 0.5Å a ≤ -0.5Å. **A**: la tripsina ultracompacta 3GY2 superpuesta sobre la tripsina 3GY4, de compactación normal. El color rojo abarca todo el diagrama porque hay una contracción general de 3GY2 hacia el CM. **B**: 3GY5 sobre 3GY2. Las dos estructuras son ultracompactas y fueron inhibidas con diminazeno. El dominio del blanco en todo el diagrama es consistente con lo observado en la figura 4.15, donde se observa que la distancia a CM es prácticamente igual en ambas estructuras, para todos los átomos de sus cadenas principales. Esto implica que el perfil de compactación es equivalente 'punto a punto'. **C**: 3GY3 sobre 3GY2. Las dos son ultracompactas, pero 3GY3 fue inhibida con pentamidina. El diagrama mayoritariamente en blanco es la consecuencia de un perfil de compactación casi idéntico de las dos estructuras comparadas. Sin embargo, la helice α en la región C– terminal en rojo indica que ese dominio se acerca más a CM en 3GY3 que en 3GY2.

4.5. β-2-microglobulina (β2M)

La proteína β2M es la cadena liviana del complejo mayor de histocompatibilidad de clase I (CMH-I). Se encuentra asociada de forma no covalente con la cadena α del CMH - I en la superficie de la mayoría de las células nucleadas, pero también se le puede hallar en plasma como proteína de dominio único, y su abundancia es un marcador diagnóstico y pronóstico en algunos desórdenes patológicos. La función principal del complejo CMH - I es la presentación de péptidos derivados de proteínas citosólicas a linfocitos T CD8+ citotóxicos, y es la diana de reconocimiento de los receptores inhibidores de las células NK ('natural killer'). Casi 500 entradas en el PDB incluyen al menos una cadena de β2M en más de 30 arquitecturas cuaternarias

distintas (InterPro : IPR015707;[43]).

En nuestro trabajo fueron analizadas 582 estructuras de β2M pertenecientes a tres clases de cadenas. Dos de éstas son de β -2- microglobulina humana (β2M de 99 aminoácidos y β2M de 100 aminoácidos) y solamente se distinguen porque en una clase se conserva el residuo de metionina inicial y en la otra no. Esta diferencia es suficiente para que hayan debido ser consideradas como clases distintas en el análisis estadístico, aún cuando corresponden a la misma proteína de la misma especie. La tercera clase reúne a las cadenas de β2M murinas.

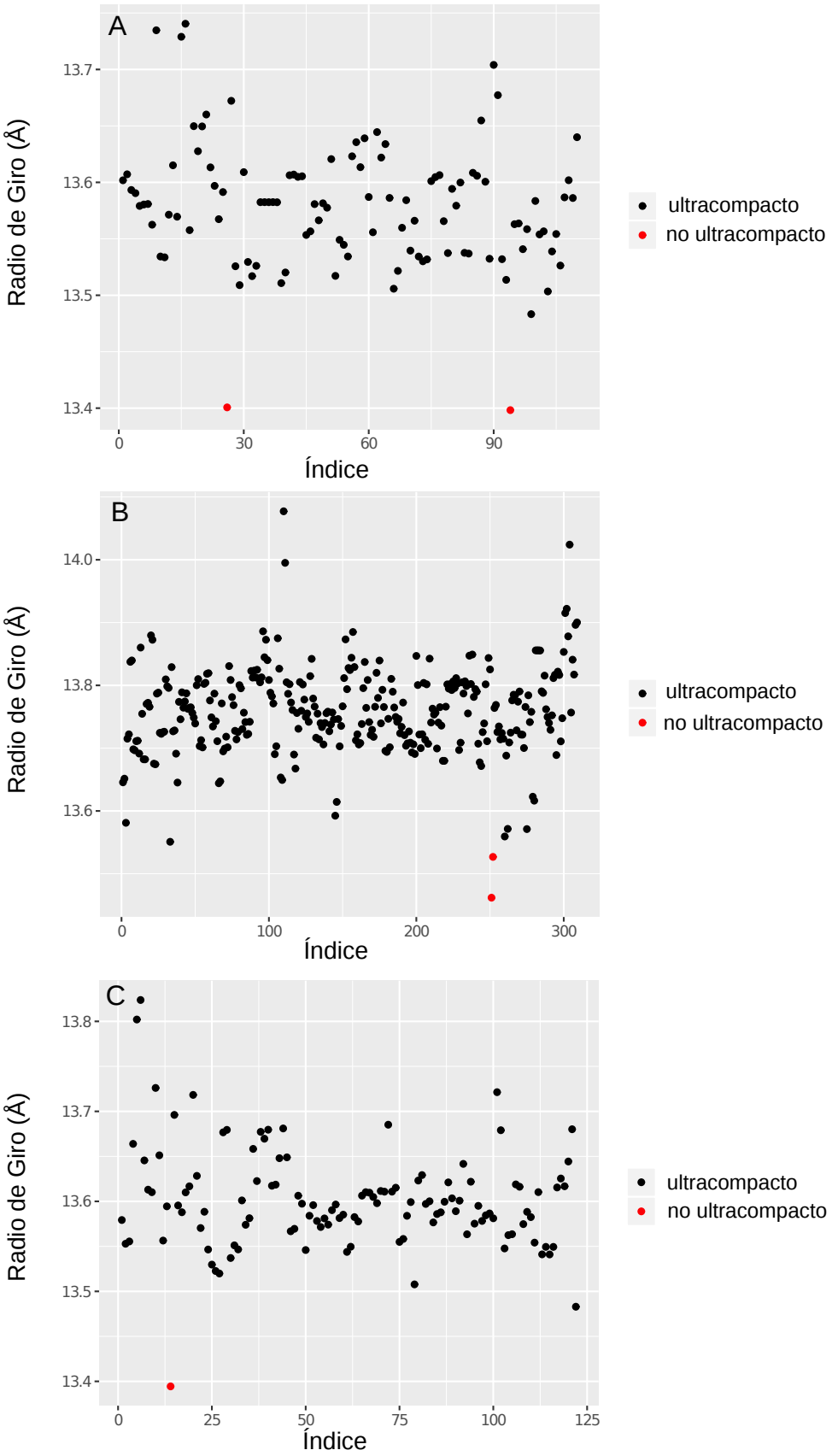


Figura 4.17: R_G **de tres clases de cadenas de** β**-2-microglobulina (β2M).** **A**: β2M humana de 99 aminoácidos. **B**: β2M humana de 100 aminoácidos (conserva la metionina inicial). **C**: β2M murina. En rojo se destacan las cadenas ultracompactas.

Entre las β2M de procedencia humana se identificaron cuatro estructuras ultracompactas (identificación de PDB: 3VFN, 1TIZ, y dos cadenas en la entrada del PDB: 3UTT). Asímismo, se halló una cadena ultracompacta de la β2M de ratón (Tabla 4.5).

PDB ID	Átomos	R_G	C—N	TCD	Valor Z 1	Rango TCD	Valor Z 2
3VFN B	396	13.40	1.32	100.00	-3.29	TCD ultrabaja	-4.31
1T1Z B	396	13.40	1.32	100.00	-3.24	TCD ultrabaja	-4.25
3UTT B	400	13.46	1.32	100.00	-4.21	TCD ultrabaja	-5.55
3UTT G	400	13.53	1.32	100.00	-3.29	TCD ultrabaja	-4.35
1K8D B	396	13.39	1.32	100.00	-3.71	TCD ultrabaja	-5.06

Tabla 4.5: **Estructuras ultracompactas de β2M.** Las primeras dos pertenecen a la β2M humana de 99 aminoácidos, las dos siguientes a las de clase humana de 100 aminoácidos y la última a la β2M de ratón. Como se aprecia en el número de átomos, tanto las dos primeras β2M humanas como la murina tiene 396 átomos distintos del hidrógeno en la cadena principal, los que corresponden a 99 aminoácidos. El R_G de las cadenas ultracompactas es muy similar en todas las clases.

El depósito 3VFN se corresponde con el heterodímero formado por la β2M con la cadena α de CMH - I que porta la mutación R151A, unido al péptido de 13 aminoácidos de virus Esptein Barr de secuencia LPEPLPQGQLTAY (HLA B*3508LPEP151A) [37]. 3VFN es una estructura reportada en el contexto de una investigación donde se resolvieron diez estructuras de HLA B*3508 con el mismo ligando, siendo la diferencia entre ellas un único aminoácido mutado en la cadena α o bien una mutación puntual en el péptido reconocido, mientras que las condiciones de cristalización, la TCD, el refinamiento y las propiedades del cristal fueron idénticas. Las nueve estructuras restantes (3VFM, 3VFO, 3VFP, 3VFR, 3VFS, 3VFT, 3VFU, 3VFV, 3VFW) contienen una subunidad β2M con R_G normal. La mutación en la cadena α, o en el antígeno unido a ella, en todos los casos está muy lejos de la subunidad β2M que se encuentra asociada al complejo de forma no covalente. El hecho de que una de las diez β2M sea ultracompacta sugiere que este fenómeno obedece a diferencias muy sutiles en la condiciones

experimentales de preparación del cristal o de la obtención del espectro de difracción. En la figura 4.18 es apreciable el desplazamiento generalizado de los átomos de la cadena principal de la β2M de 3VFN hacia el CM de la estructura.

PDB ID	Átomos	R_G (Å)	Compactación (% R_G)	Aproximación (átomos)	Factor de aproximación (%)
3VFN B	396	13.40	1.29	380	95.96
1T1Z B	396	13.40	1.28	382	96.46
3UTT B	400	13.46	2.14	306	76.5
3UTT G	400	13.53	1.67	291	72.5
1K8D B	396	13.39	1.53	349	88.13

Tabla 4.6: **Aproximación a CM en β2M ultracompactas.** Los datos corresponden a las cadenas de β2M humana y murina con R_G desviado al menos 4 σ hacia la iquierda de la distribución de su clase respectiva. La columna Átomos corresponde al número de átomos totales de la cadena principal sin hidrógenos. Compactación es el valor porcentual de acortamiento de R_G con relación a la media de la distribución. La Aproximación es la cantidad de átomos de la cadena principal de las estructuras ultracompactas que se acercan al centro de masa con respecto a la distancia media de los átomos respectivos en la clase. El Factor de aproximación es el porcentaje que representa la Aproximación de los átomos totales

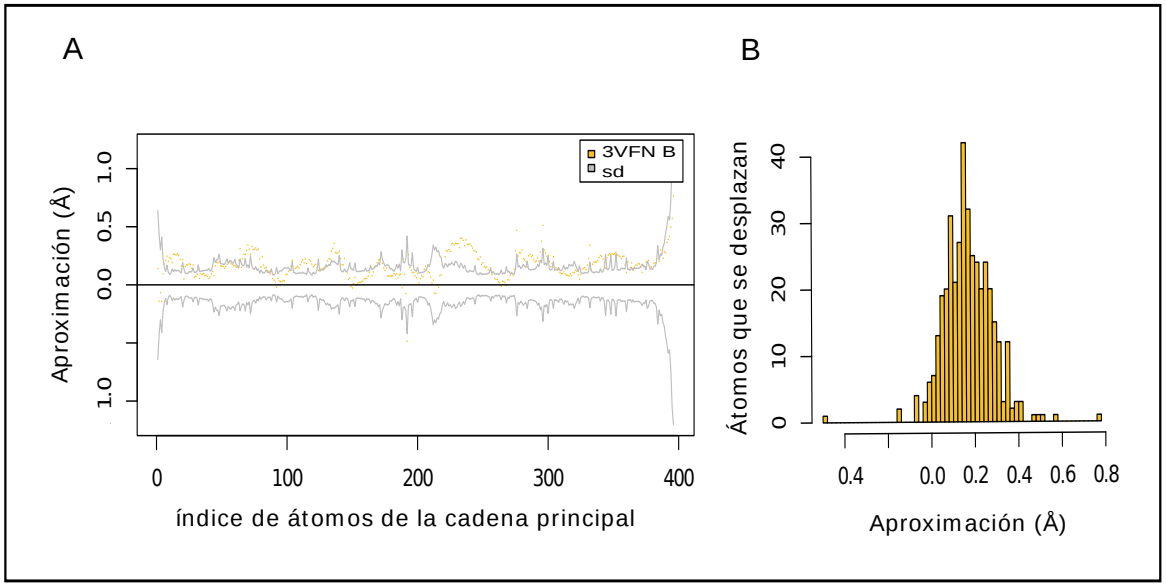


Figura 4.18: **Diferencias de distancias al CM en β2M de PDB: 3VFN.** **A**: distancias al CM de los átomos de la cadena principal para la β2M de PDB: 3VFN en Å. La mayoría de los átomos se aproximan al CM con respecto al promedio de distancia en la población de estructuras resueltas a temperatura ultrabaja. Las líneas en gris grafican ± *sd* en cada posición de la cadena principal. Valores positivos indican aproximación a CM; valores negativos indican alejamiento; y el valor cero indica conservación de la distancia. **A**: histograma de los valores de aproximación al CM con respecto al valor promedio. La contracción es evidente por la asimetría del histograma con respecto al cero, que representa la conservación de la distancia [22].

La siguente estructura ultracompacta de β2M en la Tabla 4.5 proviene de la entrada del PDB: 1T1Z. Esta β2M de 99 aminoácidos forma parte de un conjunto de ocho estructuras de HLA-A2 (CMH-I humano) unidos a mutantes de un péptido derivado de la proteína Gag del Virus de la Inmunodeficiencia Humana, reportada por el mismo grupo de investigación. Todos los parámetros experimentales para lo estudios de cristalografía de rayos X fueron idénticos. La única diferencia relevante fue que mientras el complejo entre el péptido salvaje y HLA A2 formó un cristal de tipo ortorrómbico, los siete mutantes y sus HLA A2 asociados formaron cristales monoclínicos. La β2M de 1T1Z es la única con R_G ultracompacto y su cristal pertenece a la geometría monoclínica mayoritaria.

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

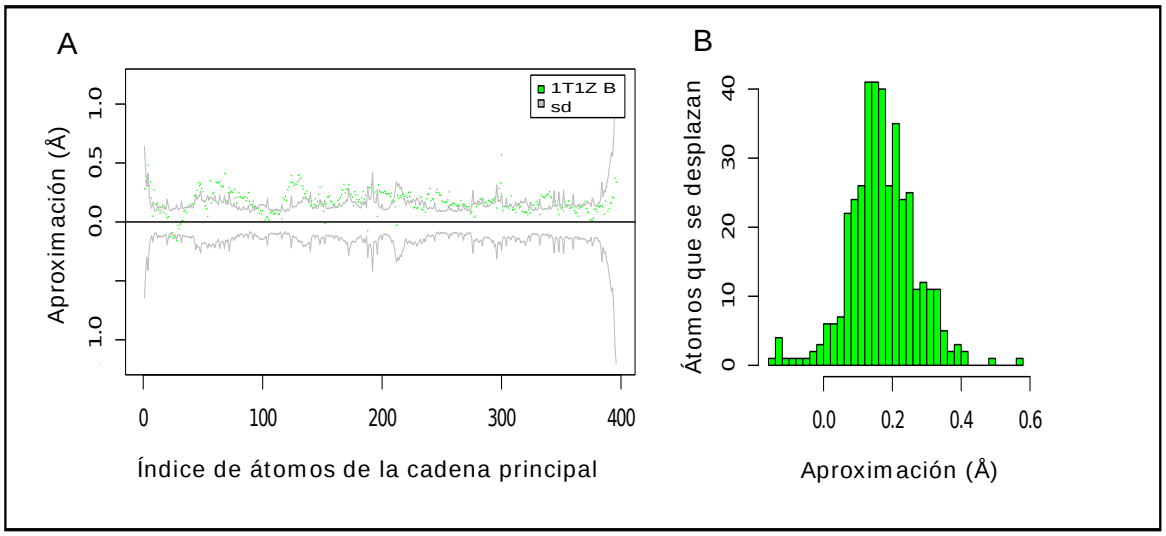


Figura 4.19: **Diferencias de distancias al CM en β2M de PDB: 1T1Z.** **A**: distancias al CM de los átomos de la cadena principal para la β2M de PDB: 1T1Z en Å. La mayoría de los átomos se aproximan al CM con respecto al promedio de distancia en la población de estructuras resueltas a temperatura ultrabaja. La asimetría del histograma con respecto al cero, que representa la conservación de la distancia al CM, confirma la compactación [22].

El gráfico de dispersión y el histograma de la aproximación al CM de los átomos de la cadena principal en la β2M de 1T1Z con respecto al promedio respectivo en la clase (figura 4.19) revelan la contracción generalizada de la estructura sobre las coordenadas centrales de la proteína (coordenadas del CM). Lo mismo sucede con las estructuras que siguen en la tabla 4.5, las cadenas B y G del depósito 3UTT del PDB. Estas últimas β2M humanas corresponden al complejo formado por un receptor de células T expresado por el clon 1E6 del linfocito T CD8+ citotóxico junto con un péptido derivado del péptido señal de la insulina humana y la HLA A2 como complejo mayor de histocompatibilidad de clase I. En estos casos, la celda unidad contiene un dímero del complejo y, en consecuencia, dos moléculas de β2M: las dos ultracompactas, como se observa en la figura 4.20.

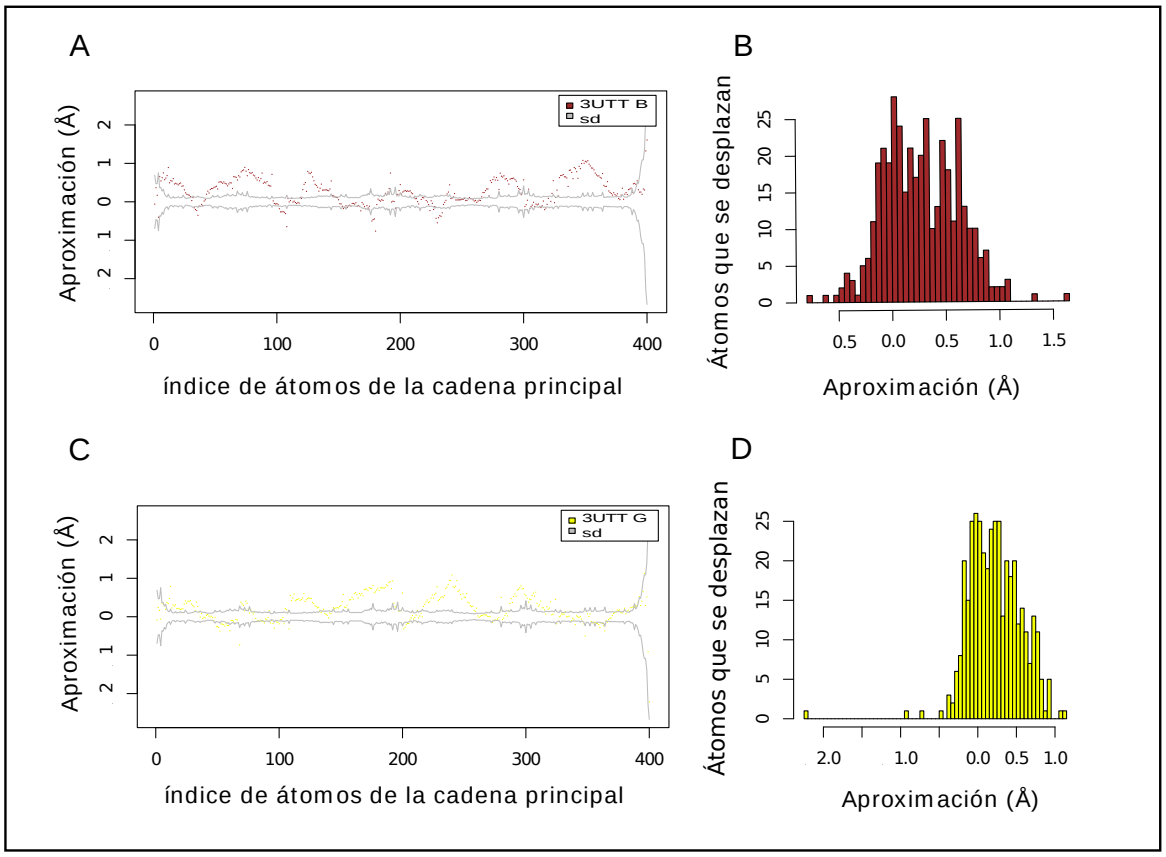


Figura 4.20: **Diferencias de distancias al CM en las β2M de PDB: 3UTT.** En **A** y **C** se muestran las distancias al CM de los átomos distintos del hidrógeno de las cadenas principales en las β2M de PDB: 3UTT— B y 3UTT — G, en Å. En los dos casos la mayoría de los átomos se aproximan al CM con respecto al promedio de distancia en la población de estructuras resueltas a temperatura ultrabaja. **B** y **D** son los histogramas de los valores de aproximación al CM con respecto al valor promedio en 3UTT — B y 3UTT — G

Basta con observar las representaciones de Richardson de la superposición con las estructuras de referencia (figura 4.21) para advertir que, más allá de la similitud global, el patrón de compactación no es el mismo en las cuatro estructuras. Esto significa que el cambio conformacional que conduce del tamaño normal hacia la ultracompactación no es una transición entre dos estados únicos bien definidos. Por el contrario, el resultado implica la existencia de formas compactas heterógeneas, tal como si cada estructura particular explorara un ensamble conformacional de estructuras ultracompactas.

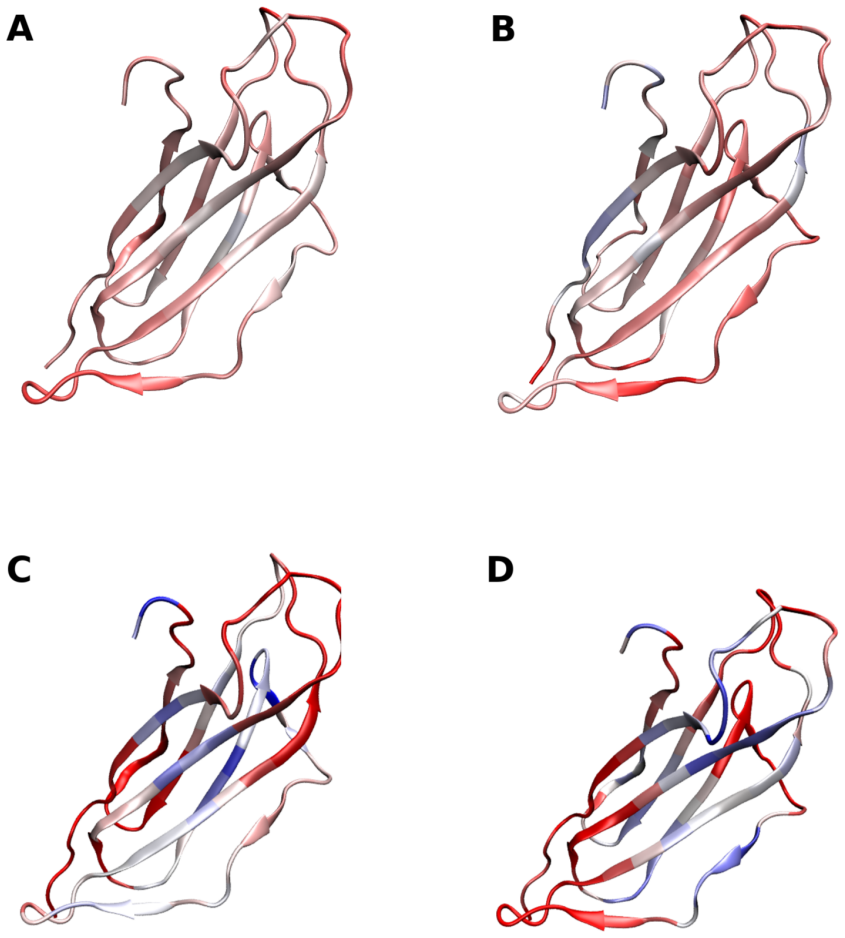


Figura 4.21: **Diagramas de Richardson de β2M humanas ultracompactas 'superpuestas' con estructuras de referencia.** La diferencia de distancia al CM se representa en la escala de colores rojo — blanco — azul en el rango ≥ 0.5Å a ≤ -0.5Å. La mayoría de los átomos de las cadenas ultracompactas están más próximos (rojo) al CM que en las cadenas de referencia, sin embargo la compactación no es tan notable ni sigue un patrón evidente como en el caso de las tripsinas.

La última estructura de la tabla 4.5 corresponde a la β2M de ratón ultracompacta con depósito en el PDB: 1K8D. En los gráficos de diferencia de distancias al CM de los átomos de su cadena principal con respecto al promedio de la clase (figura 4.22) y en la representación con escala de colores del diagrama de Richardson de la superposición con la β2M del depósito 3P4N ($Z = 0$), se oberva la compactación general (figura 4.23). En la tabla 4.6 se contabiliza el grado

de compactación: 349 átomos del esqueleto peptídico se aproximan al CM en comparación con el promedio de la distancias respectivas en la clase de las β2M murinas.

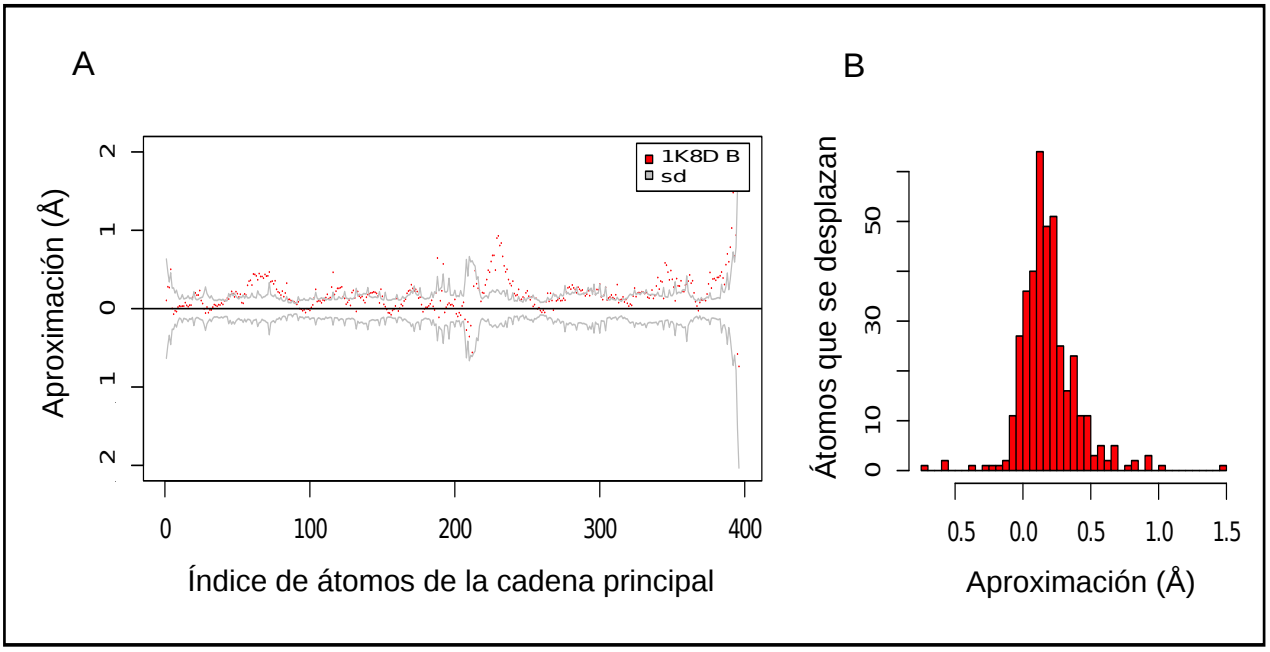


Figura 4.22: **Diferencias de distancias al centro de masa en β2M murina de PDB: 1K8D.** **A**: distancias al CMa de los átomos de la cadena principal para la β2M de PDB: 1K8D, en Å. **B**: el histograma de las diferencias de distancias revela la compactación.

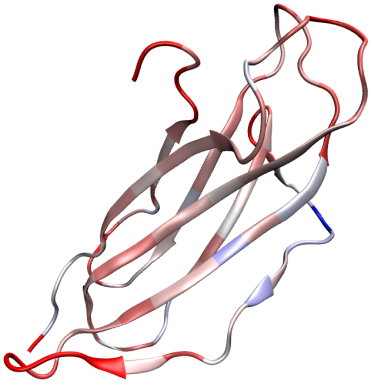


Figura 4.23: **Representación de Richardson de β2M murina ultracompacta (PDB: 1K8D) superpuesta con β2M de referencia (PDB: 3P4N).** Diferencia de distancia al CM se representa en la escala de colores rojo — blanco — azul en el rango ≥ 0.5Å a ≤ -0.5Å. La mayoría de los átomos en la cadena ultracompacta están más próximos al CM que en la referencia.

4.6. Ciclina A2

La ciclina A2 es un reguladora de quinasas dependiente de ciclinas (CDKs). Mediante la formación de complejos con CDK1 y CDK2 promueve la transición G1/S y G2/M del ciclo celular. Hay más de 90 entradas en el PDB de ciclina A2, sin embargo sólo 26 se corresponden en el fragmento C- terminal de 259 aminoácidos que va desde la Valina 173 a la Leucina 432. El resto de las ciclinas A2 están cristalizadas como fragmentos muy levemente distintos, pero basta con que difieran en la longitud de un residuo o en la identidad de un aminoácido para que no integren la misma clase (ver Materiales y Métodos). Además los rigurosos controles de calidad obligaron a descartar cinco cadenas más, por lo que el universo muestral analizado incluyó 21 cadenas de ciclina A2. Todas ellas en complejo con CDK2 y, en algunos casos, con otros ligandos (figura 4.24).

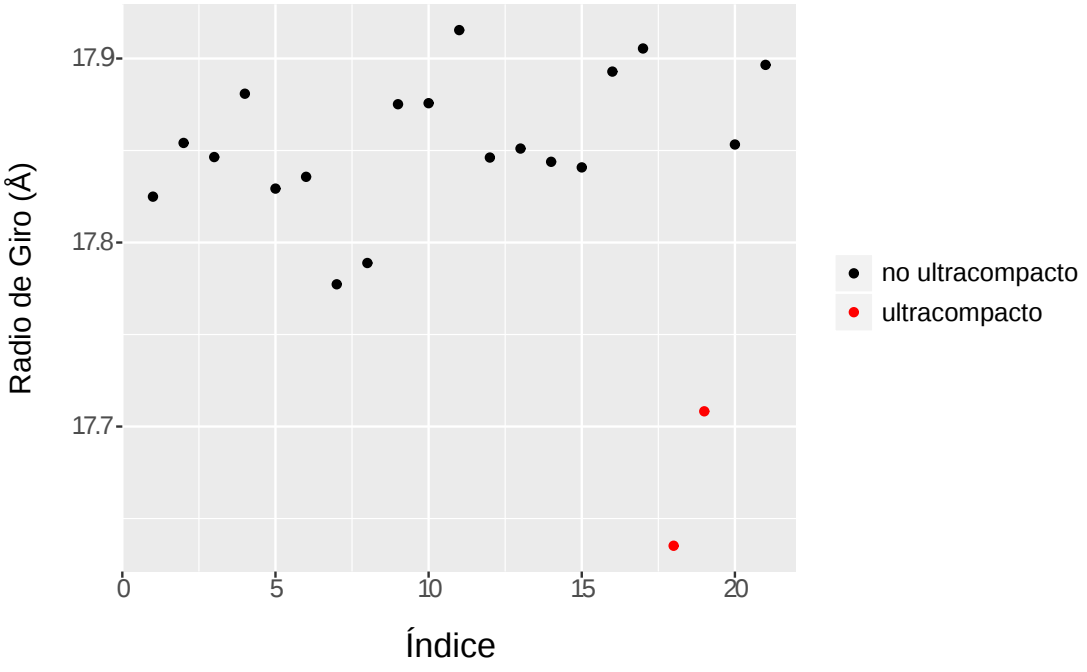


Figura 4.24: **Gráfico de dispersión de R_G de a clase de ciclina A2.** En rojo se destacan las dos estructuras cuyos R_G se encuentran en el rango de ultracompactación (glsPDB: 2WMA-B y 2WMA -D).

De las 21 cadenas de ciclina A2 analizadas, hay dos que son formas ultracompactas en virtud de sus R_G y del desplazamiento de la mayoría de los átomos de sus cadenas principales hacia las coordenadas del CM. Ambas pertenecen al mismo depósito en el PDB por lo que resultan de idénticas condiciones experimentales y están contenidas en una única unidad asimétrica. En la tabla 4.7 se detallan los parámetros de las dos cadenas ultracompactas.

PDB ID	Átomos	R_G	C—N	TCD	Valor Z 1	Rango TCD	Valor Z 2	
2WMA B	1032	17.64	1.33	100.00	-3.06	TCD ultrabaja	ultracompacto	-8.22
2WMA D	1032	17.71	1.33	100.00	-1.95	TCD ultrabaja	ultracompacto	-5.58

Tabla 4.7: **Estructuras ultracompactas de Ciclinas 2A.** La columna Átomos refiere al número de átomos distintos del hidrógeno en la cadena principal. La columna C—N corresponde a la longitud promedio de los enlaces peptídicos entre carbono y nitrógeno en Å. La existencia de dos valores Z (Z 1 y Z 2) se explica en el texto en detalle. El rango de TCD indica que TCD menores de 160 K son ultrabajas y mayor a 240 K son moderadas.

En la tabla 4.8 se listan las características de la compactación. La ciclina A2 2WMA- B muestra una contracción más generalizada que abarca más del 90 % de los átomos, mientras que 2WMA - D se compacta menos tanto en el R_G como en la cantidad de átomos que efectivamente se aproximan al CM. Al igual en β2M y a diferencia de las tripsinas del grupo 3GYs, los perfiles de ultracompactación de las ciclinas 2WMAs no obedecen un patrón que hayamos identificado (figura 4.25).

PDB ID	Átomos totales (Å)	R_G	Compactación (% R_G)	Aproximación (N de átomos)	Factor de aproximación (%)
2WMA B	1032	17.64	1.13	932	90.31
2WMA D	1032	17.71	0.72	860	83.33

Tabla 4.8: **Aproximación a CM en ciclinas A2 ultracompactas.** Los datos corresponden a las cadenas de ciclina A2 con R_G desviado al menos 4 σ hacia la iquierda de la distribución de su clase respectiva. La columna Átomos corresponde al número de átomos totales de la cadena principal sin hidrógenos. Compactación es el valor porcentual de acortamiento de R_G con relación a la media de la distribución. La Aproximación es la cantidad de átomos de la cadena principal de las estructuras ultracompactas que se acercan al CM con respecto a la distancia media de los átomos respectivos en la clase. El Factor de aproximación es el porcentaje que representa la Aproximación de los átomos totales

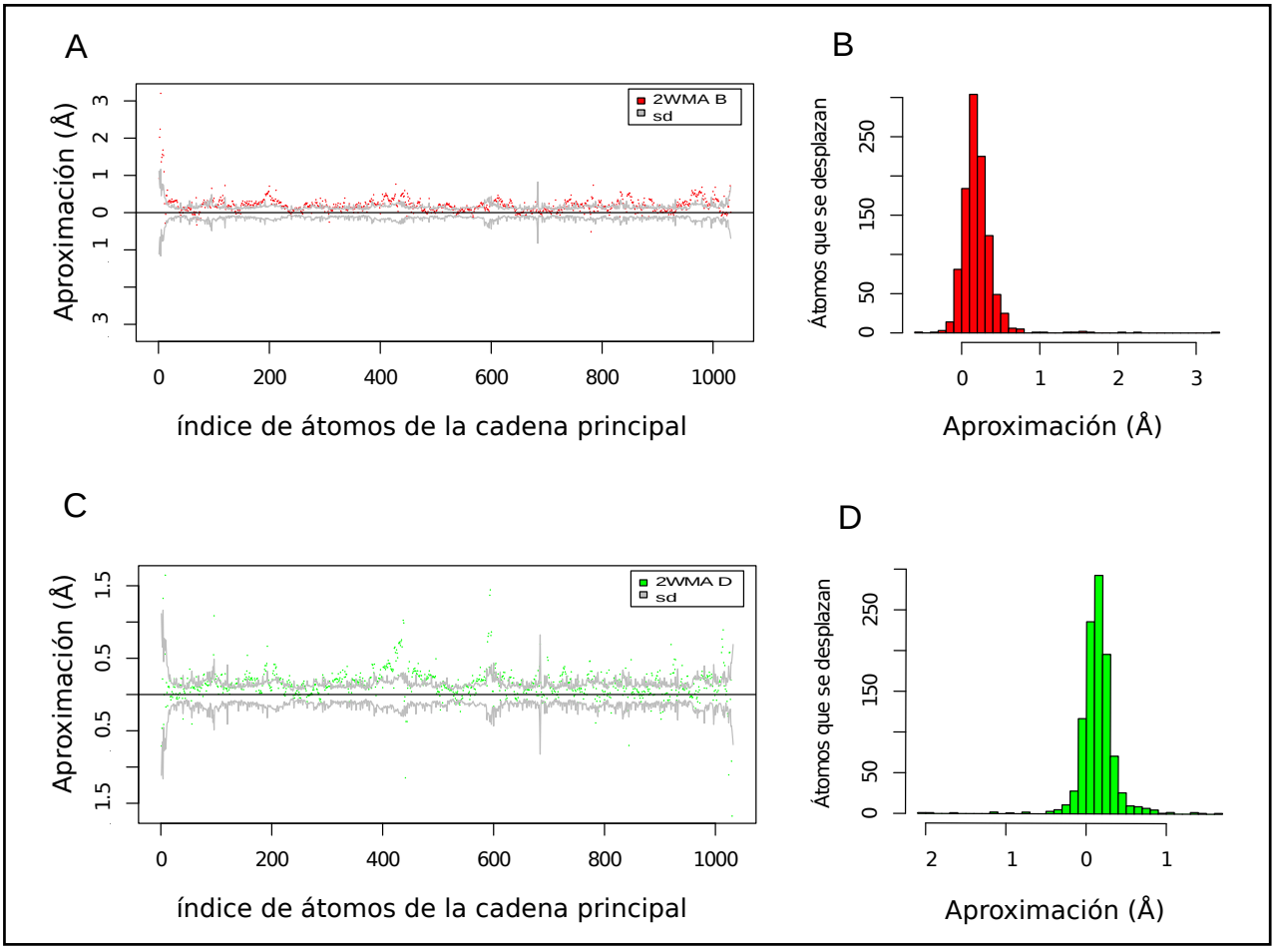


Figura 4.25: **Diferencias de distancias al CM en las ciclinas 2WMA B y 2WMA D. A**: y **C**: distancias al CM de los átomos de las cadenas principales de las ciclinas de PDB: 2WMA-B y 2WMA-D, en Å. En los dos casos, la mayoría de los átomos se aproximan al CM con respecto al promedio de distancia en la población de estructuras resueltas a temperatura ultrabaja. **B**: y **D**: histogramas de los valores de aproximación al CM con respecto al valor promedio en 2WMA-B y 2WMA-D

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

La superposición para la representación de Richardson (figura 4.26) se hizo con la ciclina A2 identificada en el PDB como 1QMZ cadena B ($Z = 0.45$). Esta estructura fue seleccionada como referencia porque tiene el valor de R_G más próximo al promedio de la distribución y comparte los parámetros experimentales más importantes. Son apreciables las diferencias en el patrón de colores de las dos cadenas ultracompactas, lo que sugiere, nuevamente, que no son productos de una transición entre el estado nativo y una forma única más condensada, sino de la exploración sobre un espacio conformacional de estructuras ultracompactas similares pero no idénticas.

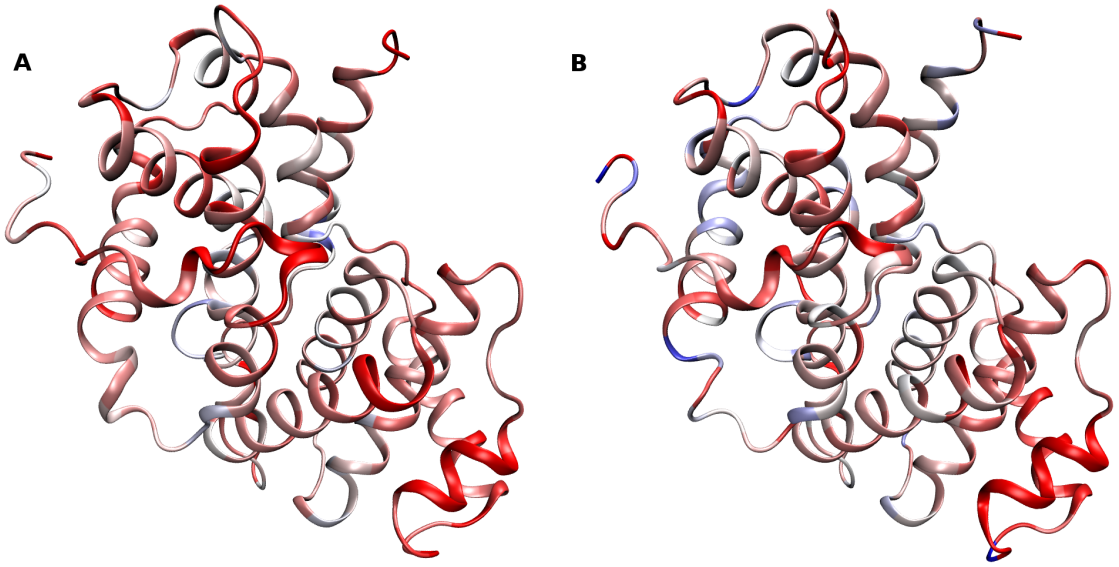


Figura 4.26: **Diagramas de Richardson de ciclinas A2 humanas ultracompactas superpuestas con la estructura de referencia 1QMZ-B.** La diferencia de distancia al CM se representa en la escala de colores rojo — blanco — azul en el rango ≥ 0.5Å a ≤ -0.5Å. La mayoría de los átomos de cadenas ultracompactas están más próximos (rojo) al CM que en la cadena de referencia, sin embargo la compactación no es tan notable ni sigue un patrón como en el caso de tripsinas.

4.7. Quimotripsina A

La clase de la cadena B de la quimotripsina A (UniProtKB P00766, residuos 16-146) se encuentra representada en el PDB por 46 estructuras cristalográficas. Del total, 24 informan

TCD y de éstas, ocho son estructuras resueltas a temperatura ultrabaja (figura 4.24)

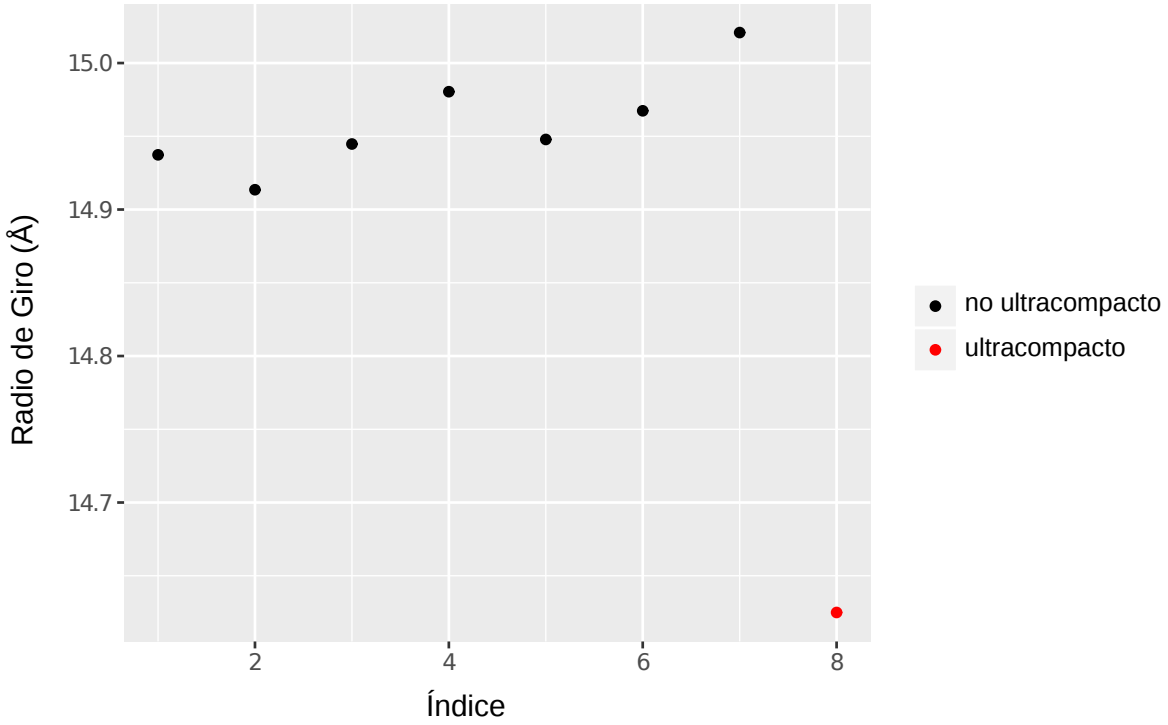


Figura 4.27: **Gráfico de dispersión de R_G en la clase de la quimotripsina A.** En rojo se destaca la estructura cuyo R_G se encuentra en el rango de ultracompactación (PDB: 3RU4-D).

.

En el análisis del set depurado se halló una estructura con valores de R_G y aproximación general de la cadena principal de la proteína al CM compatible con los criterios de ultracompactación. Esta estructura corresponde con la identificación en el PDB: 3RU4-D y exhibe un valor de $Z = -9.6$ desviaciones estándares a la izquierda de la media de la distribución de los R_G de la clase (ver tabla 4.9).

PDB ID	Átomos	R_G	C—N	TCD	Valor Z 1	Rango TCD	Compactación	Valor Z 2
3RU4 D	524	14.62	1.32	100.00	-5.17	TCD ultrabaja	ultracompacto	- 9.63

Tabla 4.9: **Estructura ultracompacta de quimotripsina A.** La columna Átomos refiere al número de átomos distintos del hidrógeno en la cadena principal. La columna C—N corresponde a la longitud promedio de los enlaces peptídicos entre carbono y nitrógeno en Å. La existencia de dos valores Z (Z 1 y Z 2) se explica en el texto en detalle. El rango de TCD indica que TCD menores de 160 K son ultrabajas y mayor a 240 K son moderadas

En la tabla 4.10 se detallan las caracerísticas de la ultracompactación. La quimotripsina 3RU4-D muestra un colapso de 2.5 % hacia el CM que involucra a prácticamente todos sus átomos (más del 99 %) sin pérdida de estructuras secundaria y terciaria, que se conservan. Un hecho notable que debe destacarse es que esta cadena de quimotripsina fue cristalizada formando un complejo triple con tripsina y el inhibidor de serina proteasas tipo Bowman-Birk, y tanto la tripsina como la quimotripsina inhibidas en el complejo presentan un R_G atípicamente bajo, dentro del criterio de ultracompactación.

PDB ID	Átomos	R_G (Å)	Compactación (% R_G)	Aproximación (átomos)	Factor de aproximación (%)
3RU4 D	524	15.00	2.49	520	99.27

Tabla 4.10: Los datos corresponden a la cadena de quimotripsina con R_G desviado al menos 4 σ hacia la iquierda de la distribución de su clase respectiva. La columna Átomos corresponde al número de átomos totales de la cadena principal sin hidrógenos. Compactación es el valor porcentual de acortamiento de R_G con relación a la media de la distribución. La Aproximación es la cantidad de átomos de la cadena principal de las estructuras ultracompactas que se acercan al centro de masa con respecto a la distancia media de los átomos respectivos en la clase. El Factor de aproximación es el porcentaje que representa la Aproximación de los átomos totales

En el gráfico de aproximación (figura 4.28) se observa con claridad el acercamiento de los átomos de la cadena principal al CM, que involucra 520 de los 524 átomos del esqueleto, mientras que los cuatro átomos restantes conservan una posición relativa similar al promedio de la distancia al CM de los átomos respectivos en el conjunto de la clase.

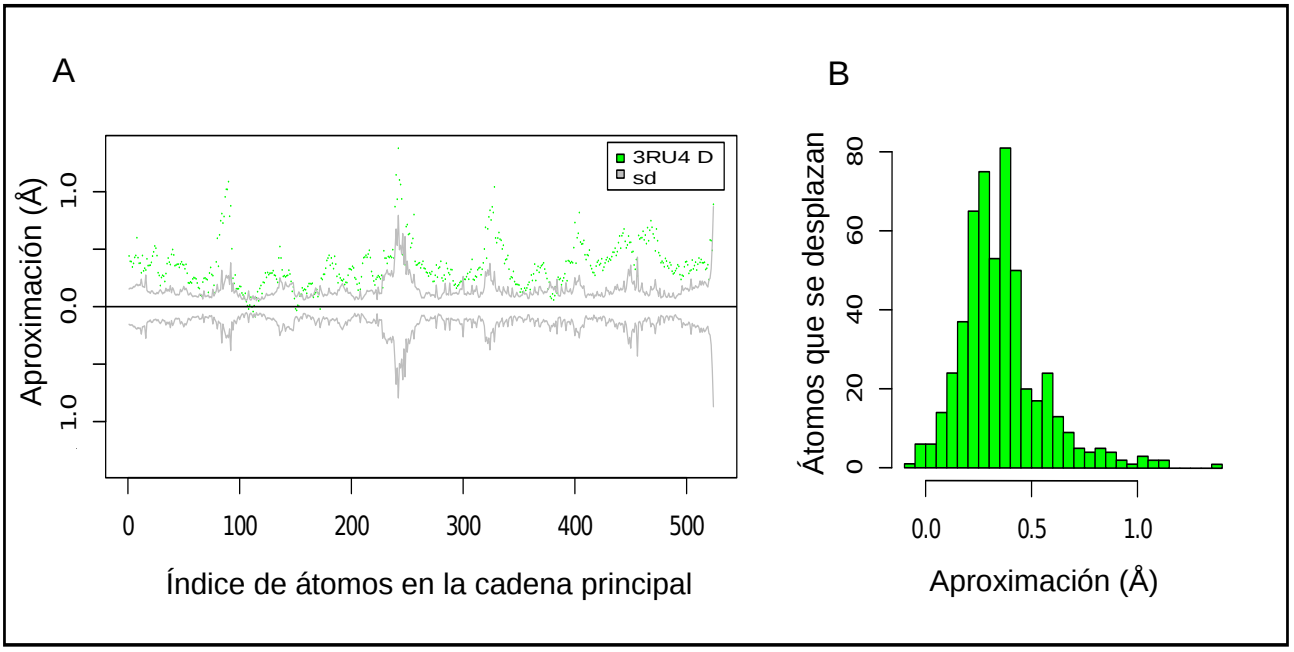


Figura 4.28: **Diferencias de distancias al CM en la quimotripsina de PDB: 3RU4-D. A**: distancias al CM de los átomos de la cadena principal de la quimotripsina 3RU4-D, en Å. La mayoría de los átomos se aproximan al CM con respecto al promedio de distancia en la población de estructuras resueltas a temperatura ultrabaja. **B**: histograma de los valores de aproximación al CM con respecto al valor promedio en 3RU4-D

La superposición para la representación de Richardson (figura 4.29) se hizo con la quimotripsina A de PDB: 1YPH-D ($Z = -0.3$). Esta estructura fue seleccionada como referencia porque tiene el valor de R_G más próximo al centro de la distribución y comparte los parámetros experimentales más importantes. Es apreciable la ultracompactación global en el patrón de colores de la figura.

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

Figura 4.29: **Representación de Richardson de quimotripsina ultracompacta 3RU4-D superpuesta con estructura de referencia 1YPH-D.** La diferencia de distancia al CM se representa en la escala de colores rojo — blanco — azul en el rango ≥ 0.5Å a ≤ -0.5Å. La mayoría de los átomos de cadenas ultracompactas están más próximos (rojo) al CM que en la cadena de referencia. La compactación es apreciable.

4.8. Otros casos: lisozima, cadena pesada de Fab, cadena liviana de Fab, cadena pesada del Factor VII de coagulación y HLA DRα

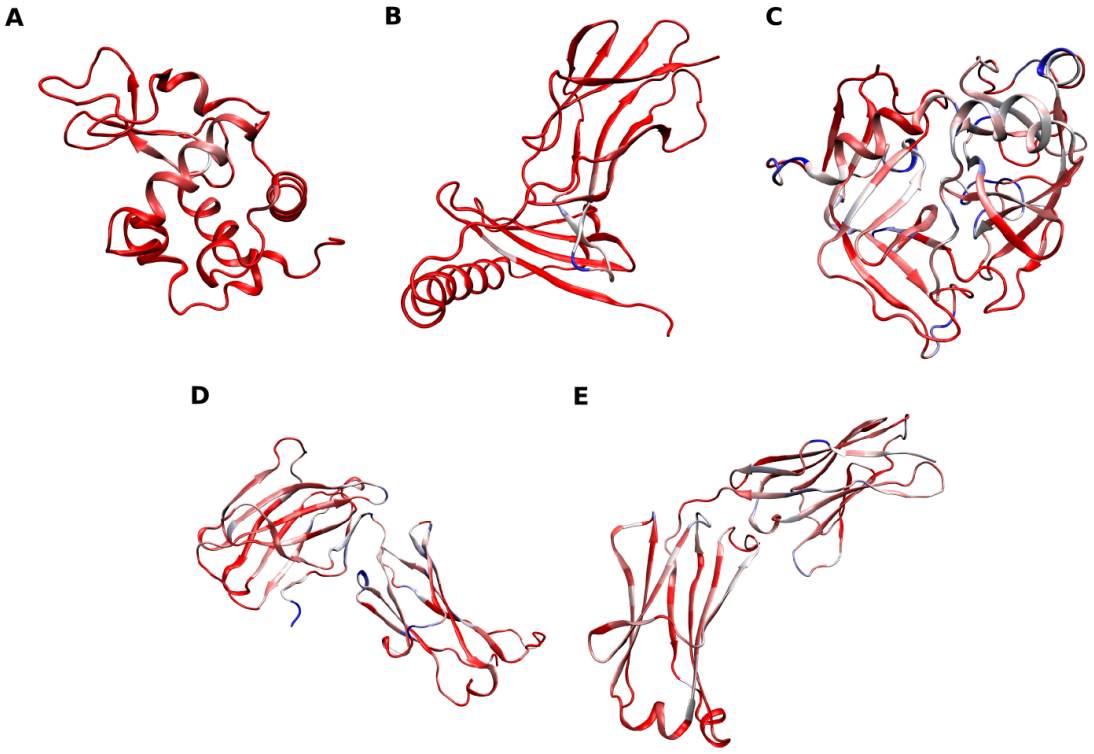


Figura 4.30: **Representaciones de Richardson de cadenas ultracompactas superpuestas con estructuras de referencia de sus clases respectivas.** La diferencia de distancia al CM se representa en la escala de colores rojo — blanco — azul en el rango ≥ 0.5Å a ≤ -0.5Å. **A**: lisozima ultracompacta 3IJV-A superpuesta sobre la lisozima 3A8Z-A, de compactación normal. El color rojo abarca todo el diagrama porque hay una contracción general hacia el CM. **B**: HLA DRα ultracompacta (2QW6 -A) sobre HLA DRα de referencia (4E41 -A). **C**: factor de coagulación VII (2FLR-H) superpuesto sobre la estructura de referencia de la clase (2CF4-H). **D**: cadena pesada de FAb ultracompacta (2DWD-A) sobre estructura de referencia de la clase (2HG5-A). **E**: cadena liviana de FAb (2DWD-B) sobre referencia de compactación típica (1ZWI-B).

Proteína	PDB ID	Átomos	R_G	C—N	TCD	Valor Z 1	Rango TCD	Compactación	Valor Z 2
Lisozima	3IJV A	516	13.26	1.32	100	-5.95	TCD ultrabaja	ultracompacto	-10.66
HLA DRa	2Q6W A	716	17.68	1.32	98	-3.10	TCD ultrabaja	ultracompacto	-14.85
CFVII	2FLR H	1016	16.42	1.30	130	-3.75	TCD ultrabaja	ultracompacto	-6.41
FAb cad pesada	2DWD A	876	20.57	1.33	100	-3.75	TCD ultrabaja	ultracompacto	-4.86
Fab cad liviana	2DWD B	848	23.018	1.33	100	-3.88	TCD ultrabaja	ultracompacto	-5.54

Tabla 4.11: **Otros casos de estructuras ultracompactas.** La columna Átomos refiere al número de átomos distintos del hidrógeno en la cadena principal. La columna C—N corresponde a la longitud promedio de los enlaces peptídicos entre carbono y nitrógeno en Å. La existencia de dos valores Z (Z 1 y Z 2) se explica en el texto en detalle. El rango de TCD indica que TCD menores de 160 K son ultrabajas y mayor a 240 K son moderadas

Proteína	PDB ID	Átomos	R_G (Å)	Compactación (% R_G)	Aproximación (átomos)	Factor de aproximación (%)
Lisozima	3IJV A	516	13.26	3.19	516	100
HLA DRa	2Q6W A	716	17.68	3.16	608	84.92
CFVII	2FLR H	1016	16.42	1.44	855	84.15
FAb cad pesada	2DWD A	876	20.57	0.88	774	88.36
FAb cad liviana	2DWD B	848	23.018	0.80	818	96.46

Tabla 4.12: **Aproximación a CM en casos ultracompactos (miscelánea).** Los datos corresponden a cadenas de diferentes clases con R_G desviado al menos 4 σ hacia la iquierda de la distribución. La columna Átomos corresponde al número de átomos totales de la cadena principal sin hidrógenos. Compactación es el valor porcentual de acortamiento de R_G con relación a la media de la distribución. La Aproximación es la cantidad de átomos de la cadena principal de las estructuras ultracompactas que se acercan al centro de masa con respecto a la distancia media de los átomos respectivos en la clase. El Factor de aproximación es el porcentaje que representa la Aproximación de los átomos totales

4.8.1. Lisozima (PDB: 3IJV cadena A)

La clase de la lisozima de la clara de huevo (UnitProtKB: P00698) tiene 221 representaciones cristalográficas en el PDB cuyas datos difracción se recolectaron a temperatura ultrabaja. Una sola de estas estructuras (3IJV-A) cumple los criterios de ultracompactación, alcanzando un valor de $Z = -10.7$ desviaciones estándares a la izquierda del promedio de la distribución de R_G de su clase y un desplazamiento relativo general de los átomos de la cadena principal hacia el CM.

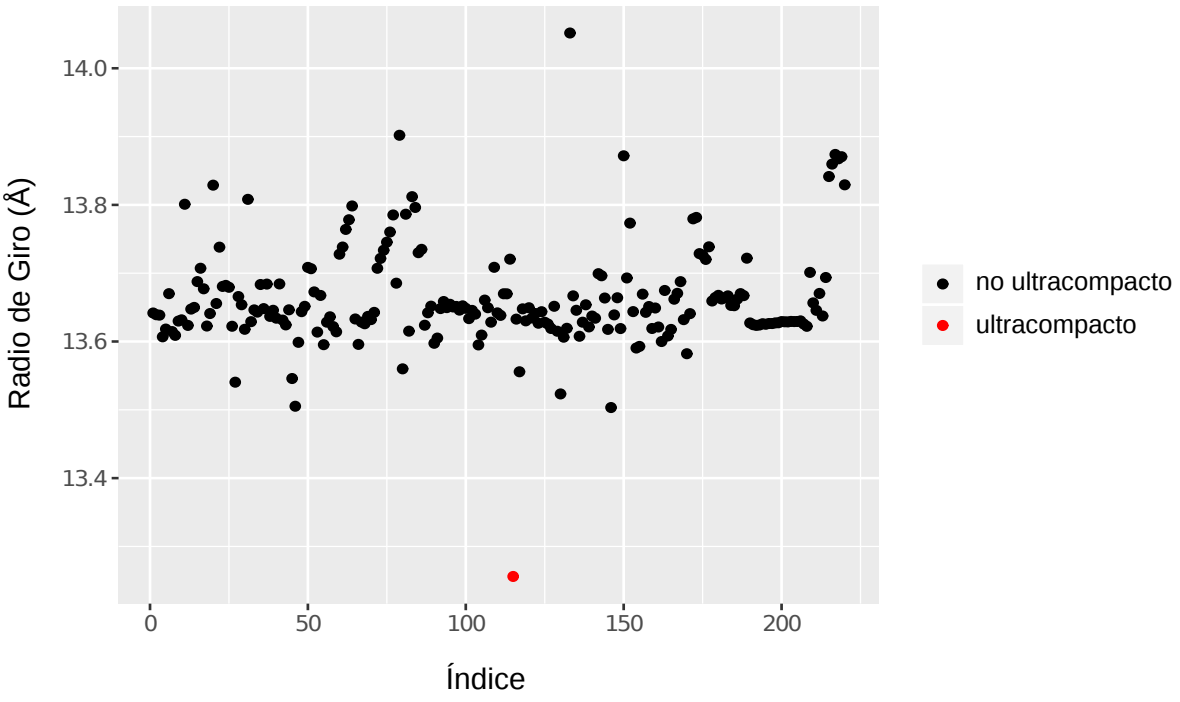


Figura 4.31: **Gráfico de dispersión de R_G en la clase de las lisozimas resueltas a temperatura ultrabaja.** En rojo se destaca la estructura cuyo R_G se encuentra en el rango de ultracompactación (3IJV-A).

La lisozima ultracompacta detectada fue resuelta en un trabajo de 2015 a cargo de Pechkova et al. [52] donde se comparó el cristal obtenido mediante un procedimiento de cristalización inducida por una plantilla de alúmina anódica nanoporosa (APA - anodic porose alumina template) con el cristal obtenido por el método clásico de difusión de vapor por gota colgante, siendo la cadena depositada en el PDB con la identificación 3IJV -A la lisozima de referencia cristalizada por la técnica tradicional. Los parámetros experimentales de 3IJV-A son prácticamente idénticos al resto de las estructuras cristalizadas por difusión de vapor, pero su cáracter ultracompacto es notable tanto en la desviación de su R_G cuanto en la distancia 'punto a punto' al CM de los átomos de su cadena principal. En las figuras 4.31 y 4.32 se observan la posición en el gráfico de dispersión de los R_G de su clase y la aproximación, que involucra al 100 % de los átomos considerados, hacia el centro de la proteína (CM).

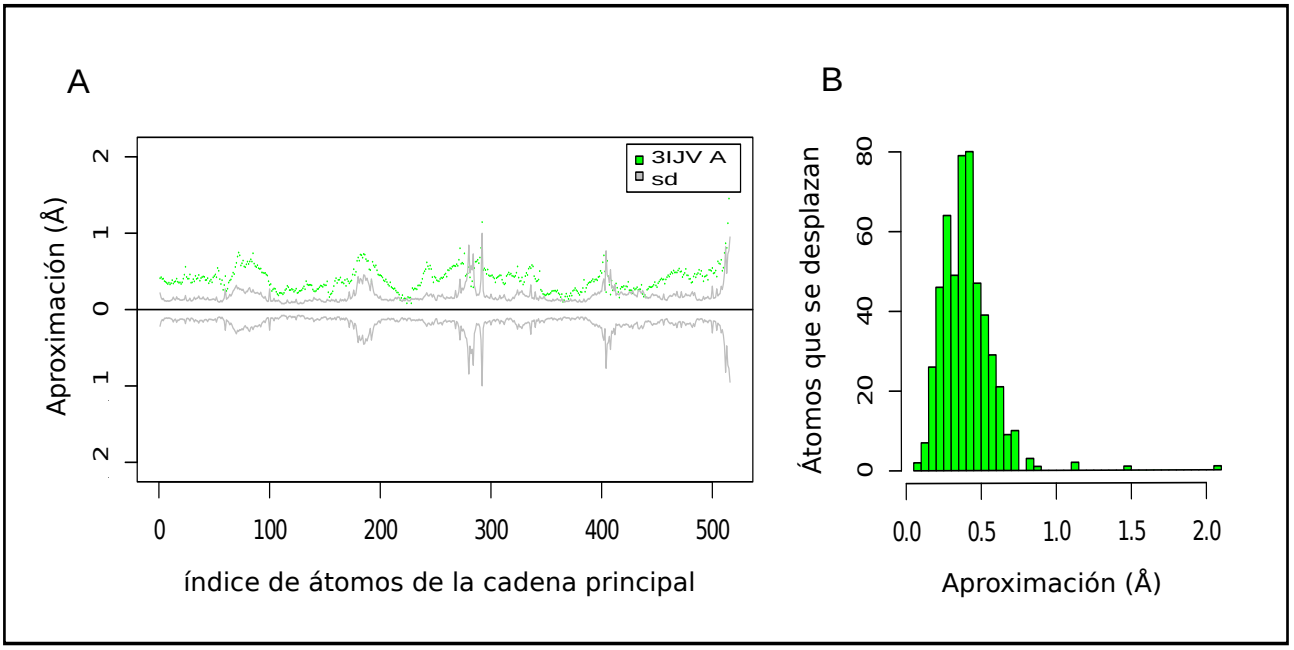


Figura 4.32: **Diferencias de distancias al CM en la lisozima 3JIV-A. A**: distancias al CM de los átomos de la cadena principal de la lisozima 3JIV-A, en Å. La totalidad de los átomos de la cadena principal se aproximan al CM con respecto al promedio de distancias a CM en la población de estructuras resueltas a temperatura ultrabaja. **B**: histograma de los valores de aproximación al CM con respecto al valor promedio en 3JIV-A

En la representación de Richardson de la 'superposión' de 3IJV-A y una estructura de referencia de R_G normal (PDB: 3A8Z cadena A) con valor $Z = 0.0$ se apecia nítidamente la compactación global de la lisozima, pese a que los parámetros experimentales así como el método de cristalización no importan diferencias metodológicas significativas que justifiquen este comportamiento 4.30.

4.8.2. HLA-DRα (PDB: 2QW6 cadena A)

El antígeno leucocitario humano HLA-DR es un receptor de superficie del complejo mayor de histocompatibilidad de clase II que participa en la presentación por la vía exógena de fragmentos peptídicos a los recepectores de los linfocitos T colaboradores. El HLA-DR es un heterodímero formado por una cadena α y una cadena β en complejo con el antígeno procesado para su

presentación. Luego de todos los pasos de control, la clase de cadena HLA-Drα (UnitProtKB: P01903) tiene 20 representantes en el PDB con total identidad de secuencia, pero son fragmentos que varían entre 178, 179 y 180 aminoácidos, por lo que deben considerarse en tres grupos de longitud diferente. El grupo de HLA-DRα con más representaciones en el PDB corresponde al fragmento de 179 aminoácidos y reúne 12 de las 20 cadenas de la clase. Nuestro estudio se hizo sobre las estructuras con esta longitud y sólo una cadena ultracompacta fue hallada, pero con la peculiaridad de que es ultracompacta incluso si se consideran las 20 cadenas de HLA-Drα, con independencia del tamaño de la estructura primaria de los fragmentos. Debido a ello, este caso se comunica a pesar de que no satisface el mínimo de 20 cadenas de secuencia y longitud idéntica que requeríamos para incorporarlo en el análisis.

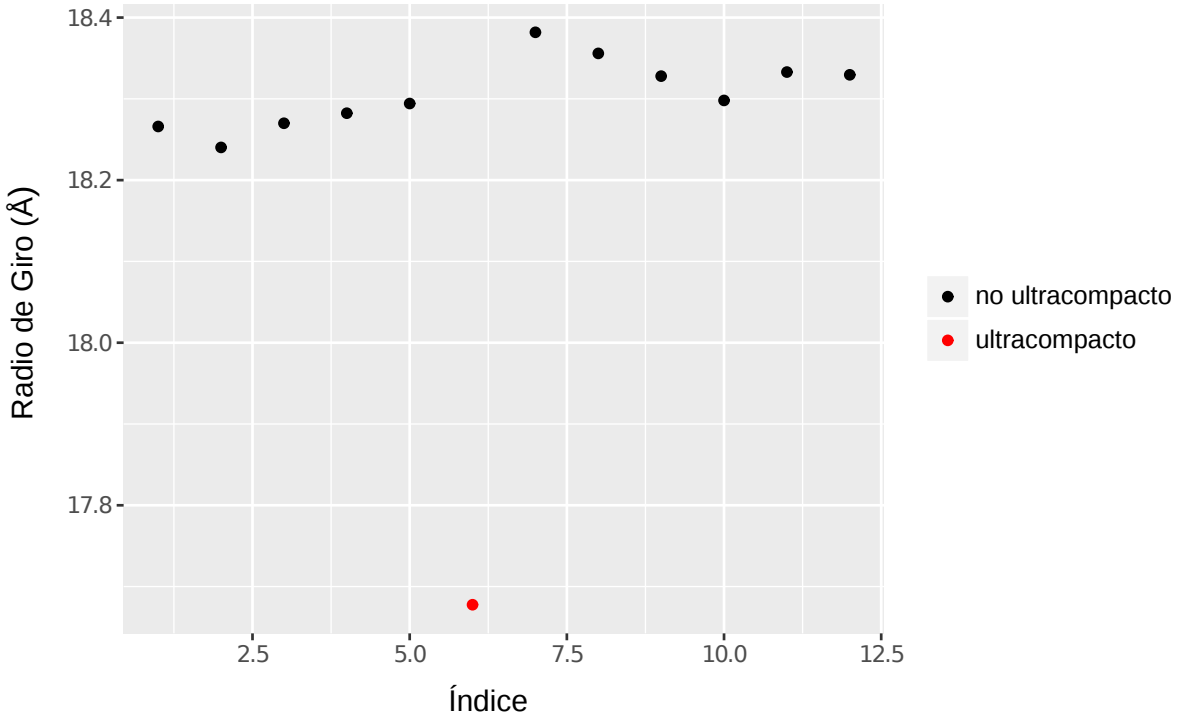


Figura 4.33: **Gráfico de dispersión de los R_G en la clase de las HLA DRα resueltas a temperatura ultrabaja.** En rojo se destaca la estructura cuyo R_G se encuentra en el rango de ultracompactación (2Q6W-A).

Las 12 cadenas de HLA-DRα de 179 aminoácidos y 716 átomos en la cadena principal fueron

resueltas con datos recolectados a temperatura ultrabaja ($\leq$ 100 K). Una de ellas, la estructura de identidad 2Q6W-A, es ultracompacta porque exhibe un valor $Z = -14.85$ a la izquierda de la media de la distribución de los R_G y, además, el 85 % de los átomos de su esqueleto peptídico se aproximan al CM en la comparación 'punto a punto' con los átomos correspondientes en el resto de las proteínas de la clase (ver figuras 4.33 y 4.34).

Como referencia para la representación de Richardson de la 'superposición' se utilizó la estructura 4E41-A de valor $Z = -0.2$. Este Z la ubica como la más próxima al valor promedio de R_G dentro de la clase. Se destaca que tanto en la celda unitaria del cristal de 2QW6 como en la referencia 4E41 hay dos cadenas HLA DRα. La estructura HLA DRα contenida en la cadena D de 2Q6W no fue incluida en el análisis porque fue depurada en un paso previo de control de calidad de las cadenas de la clase, debido a que le faltaban cuatro átomos en el equeleto peptídico. Sin embargo, el $R_G = 17.88$ Å de esta estructura descartada también era sensiblemente más pequeño que el promedio de la clase.

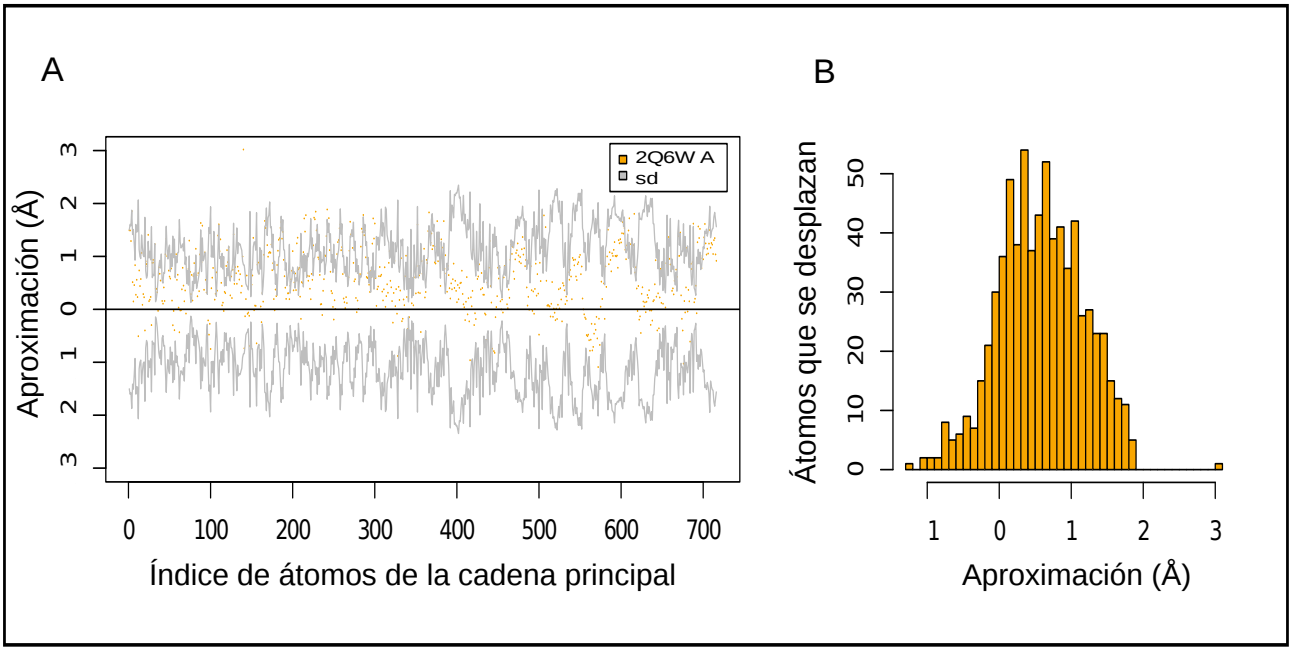


Figura 4.34: **Diferencias de distancias al CM en HLA DRα 2Q6W- A. A**: distancias al CM de los átomos de la cadena principal de la HLA DRα 2Q6W-A, en Å. La mayoría de los átomos se aproximan al CM con respecto al promedio de distancia en la población de estructuras resueltas a temperatura ultrabaja. **B**: histograma de los valores de aproximación al CM en 2Q6W-A con respecto al valor promedio en la clase.

4.8.3. Cadena pesada de Factor de Coagulación VII (PDB: 2FLR cadena H)

El factor VII de coagulación es una serina proteasa que inicia la vía extrínseca de la coagulación sanguínea. Esta proteína plasmática circula en la sangre como zimógeno y, una vez procesada, exhibe una cadena liviana y una cadena pesada. La clase de las cadenas pesadas del factor VII comprende 37 estructuras resueltas por cristalografía de rayos X, para 32 de las cuales se utilizaron datos de difracción recolectados a temperatura ultrabaja.

El análisis de esta clase permitió detectar una sola estructura ultracompacta, de acuerdo al R_G y a los valores de acercamiento relativo de los átomos de la cadena principal al CM. La estructura con identificador PDB 2FLR cadena H presenta un valor $Z = -6.41$ desviaciones hacia la izquierda del promedio de la clase y cerca del 85 % de los átomos del esqueleto peptídico

se encuentran más próximos al CM que el promedio de los átomos correspondientes en el resto de las cadenas.

En la representación de Richardson de la 'superposición' de 2FLR-H con la cadena H de 2CF4 ($Z = 0.1$) es apreciable la compactación general, codificada en color rojo, aunque no es tan notable como en los casos de otras proteínas analizadas.

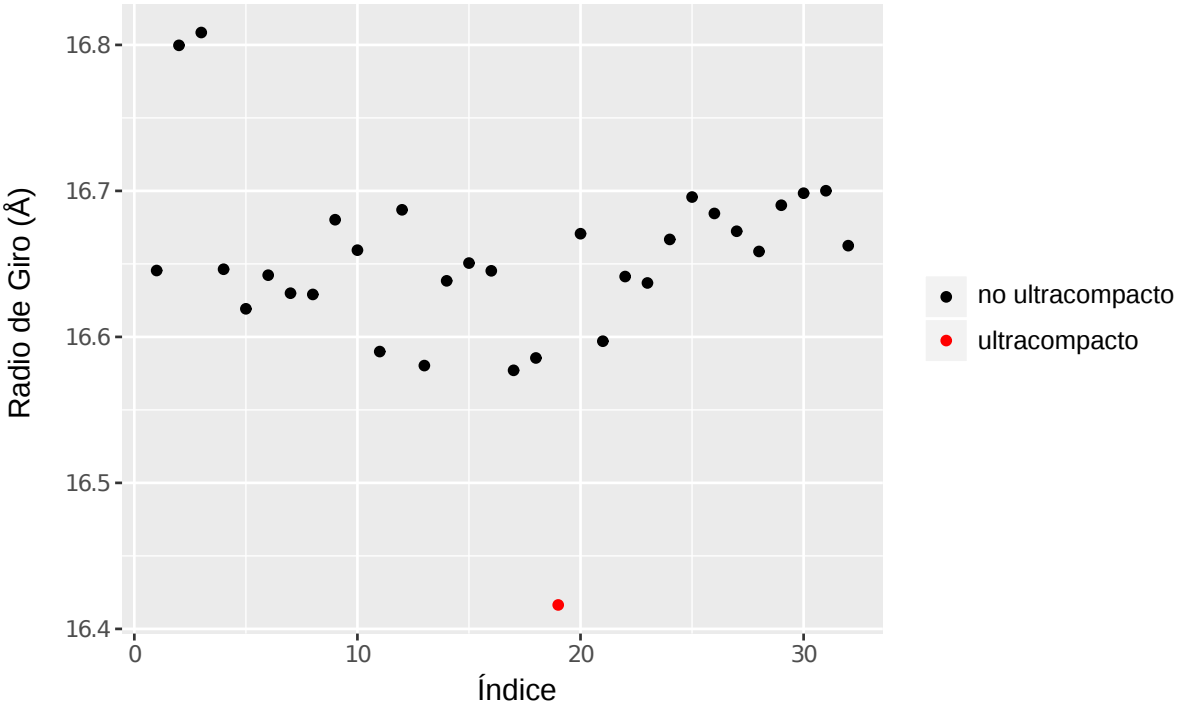


Figura 4.35: **Gráfico de dispersión de los R_G de los factores de coagulación VII resueltos a temperatura ultrabaja.** En rojo se destaca la estructura cuyo R_G se encuentra en el rango de ultracompactación (PDB: 2FLR -H).

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

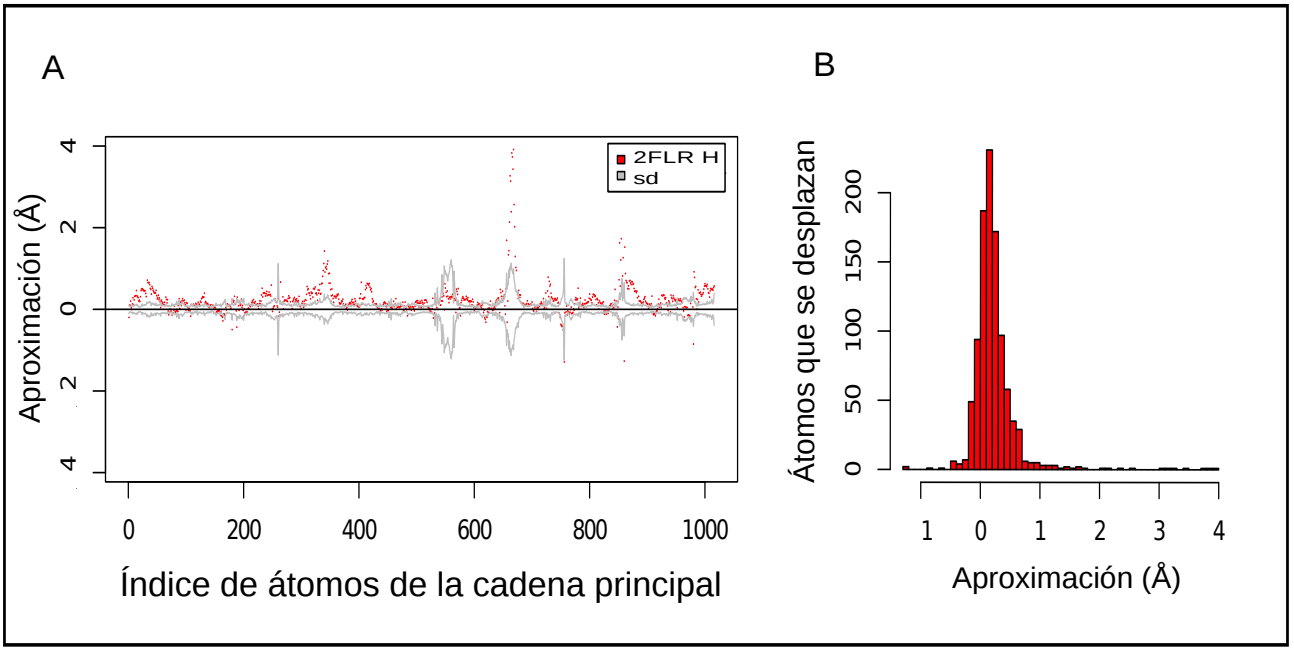


Figura 4.36: **Diferencias de distancias al CM en el factor de coagulación VII de PDB: 2FLR-H.** **A**: distancias al CM de los átomos de la cadena principal del factor de coagulación VII 2FLR-H, en Å. La mayoría de los átomos se aproximan al CM con respecto al promedio de distancia en la población de estructuras resueltas a temperatura ultrabaja. **B**: histograma de los valores de aproximación al CM con respecto al valor promedio en 2FLR-H

4.8.4. Fragmentos Fab de anticuerpo monoclonal anti KcsA

Las cadenas pesadas y las cadenas livianas de Fab (fragmento de unión al antígeno) del anticuperpo de ratón específico contra el canal de potasio KcsA conforman dos clases que comprenden 32 estructuras (cadenas) depositadas en elPDB por cada clase. En la entrada del PDB 2DWD tanto la cadena pesada (2DWD-A) como la cadena liviana (2DWD-B) satisfacen los criterios de ultracompactación según sus R_G y las distancias de los átomos de sus esqueletos peptídicos a los respectivos CM. Mientras que la cadena pesada tiene un valor de $Z = -4.9$ desviaciones estándares hacia la izquierda del promedio, la cadena liviana 2DWD-B está desplazada -5.5 desviaciones hacia los valores de compactación. Asimismo, alrededor de un 90 % de los átomos de las cadenas principales de dichas estructuras se aproximan al CM en relación

a las distancias promedios de los átomos correspondientes en sus respectivas clases.

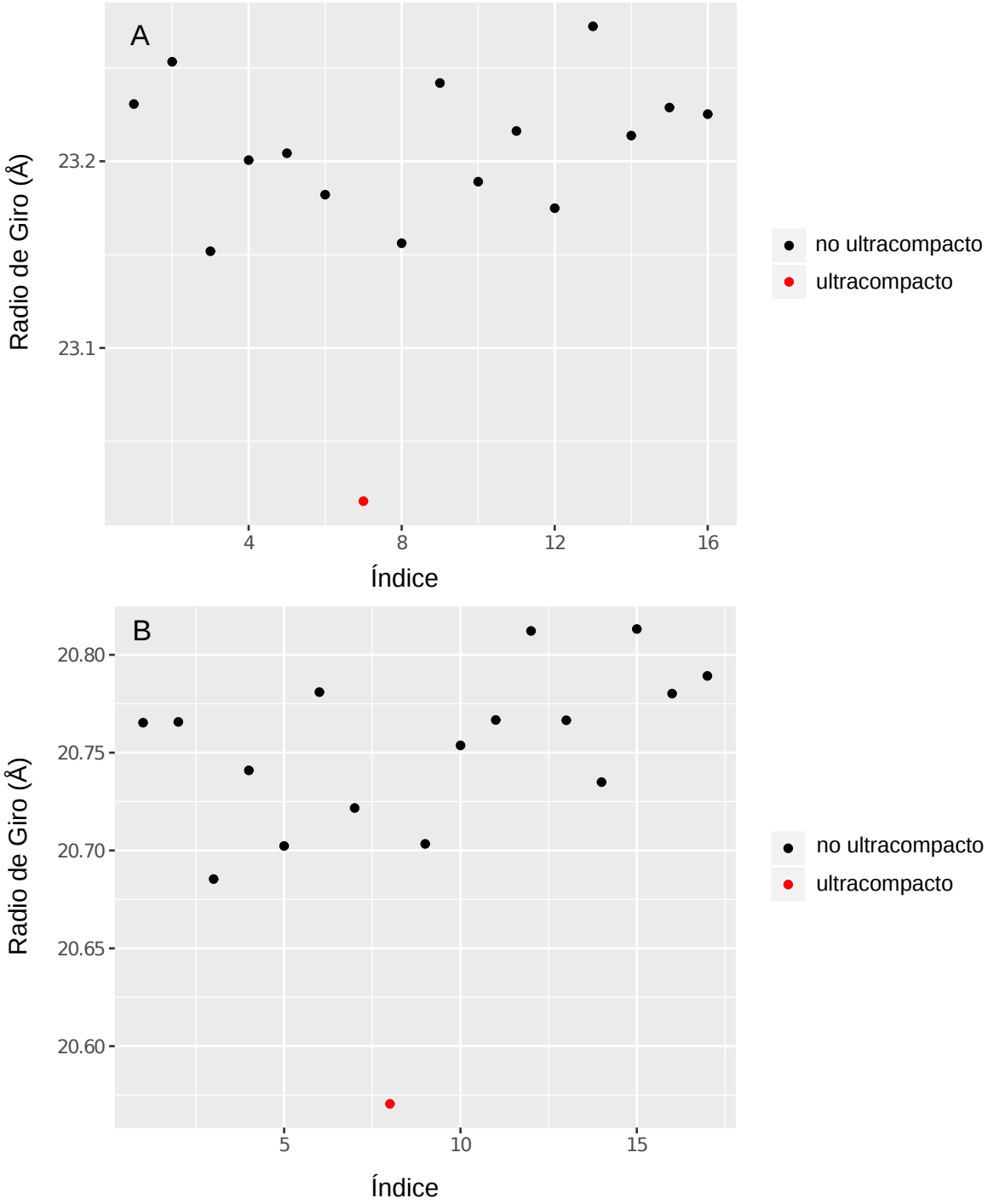


Figura 4.37: **Gráficos de dispersión de las clases del Fab anti KscA: cadena liviana y pesada.** **A**: cadena liviana de Fab anti KscA. **B**:cadena pesada de Fab anti KscA. En rojo se destacan las formas ultracompactas

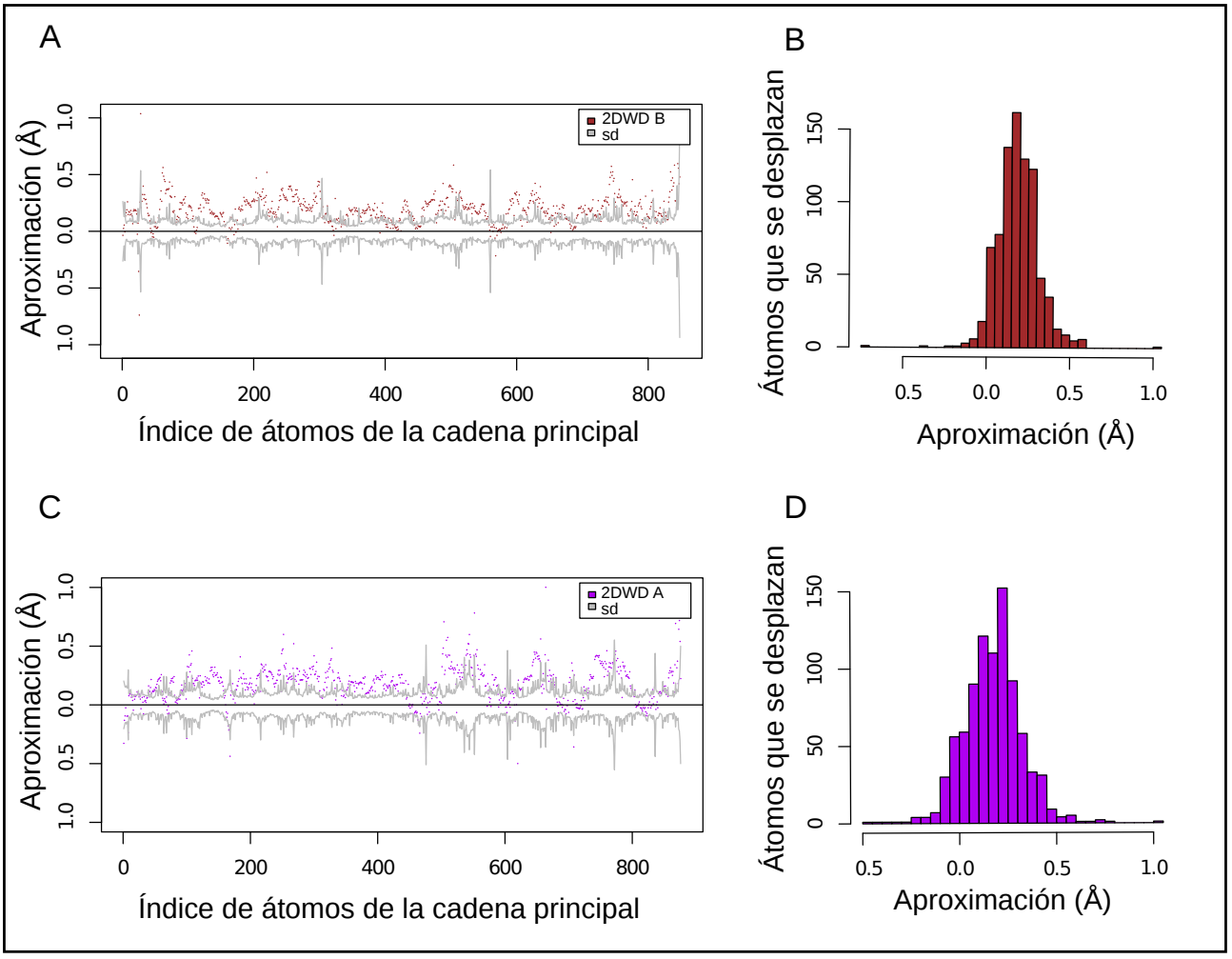


Figura 4.38: **Diferencias de distancias al CM de los átomos de las cadenas principales de 2DWD-B y 2DWD-A A** y **C**: aproximación al CM de los átomos pesados de los esqueletos peptídicos en las cadenas liviana y pesada de Fab anti KscA (2DWD-B y 2DWD-A, respectivamente). **B** y **D**: histogramas de las distancias de aproximación

Las representaciones de Richardson de las 'superposiones' fueron construidas sobre las referencias 1ZWI -B, para la cadena liviana del Fab, y 2HG5-A para la cadena pesada. Un hallazgo a resaltar es que la proteína de membrana KcsA de la entrada del PDB 2DWD también parece ser compacta, lo que nuevamente sugiere que el fenómeno afecta a toda la celda unidad y, en este caso, abre la interrogante sobre si podría existir una relación de causalidad entre el carácter compacto de la proteína diana y el carácter ultracompacto de los fragmentos Fab de unión a ella. En otros términos: si la compactación podría ser un fenómeno inducible por ligando o la

misma refleja sutiles cambios en la condiciones metodológicas de la resolución de las estructuras cristalográficas por el patrón de difracción de los rayos X.

Discusión

Capítulo 5

Discusión

5.1. Discusión

En este trabajo determinamos el (R_G) y la distancia al (CM) de los átomos de la cadena principal en 19,393 estructuras cristalográficas de proteínas depositadas en el PDB. Dos hallazgos alcanzados en este contexto merecen ser destacados: encontramos que el R_G de las cadenas cristalizadas es sensible a la (TCD) de los patrones de difracción, siendo el valor del R_G más pequeño, de forma estadísticamente significativa, cuando la recolección se hace a temperaturas criogénicas que cuando se realiza a temperatura moderada o ambiente. Adicionalmente, detectamos la existencia de casos que exhiben una compactación –menor R_G y aproximación general hacia CM– más importante de la que cabría esperar debido al efecto térmico. Estos casos no obedecen a errores o sesgos metodológicos apreciables en el procedimiento de resolución. Tampoco se pudo atribuír este comportamiento a las condiciones experimentales, a ajustes instrumentales en la recolección de los datos de disfracción de rayos X, al protocolo de crecimiento de los cristales, a los métodos de cristalización, al refinamiento de los datos obtenidos ni a la resolución alcanzada. Otras posibles explicaciones triviales fueros descartadas mediante un proceso riguroso de control de la calidad de los datos, que fue aplicado a todas las proteínas y clases de proteínas incluídas en la investigación.

Para el análisis estadístico, las estructuras cristalográficas de las cadenas de proteínas fueron agrupadas en clases de cadenas idénticas: sin mutaciones en la secuencia y de igual longitud. El R_G se computó utilizando solamente las coordenadas espaciales de los átomos de la cadena principal (esqueleto peptídico) sin los átomos de hidrógeno. Esta metodología de análisis excluye cualquier relación del fenómeno investigado con eventuales disimilitudes de las estructuras primarias o con discrepancias en la selección de los átomos para calcular el R_G. Cabe subrayar que algunas de las estructuras ultracompactas identificadas fueron resueltas en el contexto de experimentos en paralelo llevados a cabo por un mismo grupo de investigación en condiciones experimentales similares, registrándose, sin embargo, valores de R_G normales en el resto de las estructuras depositadas en el PDB.

Las clases analizadas contienen cadenas de proteínas químicamente idénticas, en estado monomérico o asociadas de forma no covalente a otras cadenas en diversas arquitecturas cuaternarias. Asímismo, en algunos casos las cadenas se encuentran formando complejos o inhibidas por un ligando, asociadas con iones y grupos prostéticos o aditivos. En la misma clase puede haber cadenas cristalizadas en condiciones químicas diferentes, pero éstas no bastan para explicar la ultracompactación, toda vez que más de una cadena dentro de una clase pueden mostrar ultracompactación sin participar del mismo patrón de interacciones.

No fue posible precisar con certeza si la compactación y la ultracompactación afectan a todas las estructuras que se encuentran en la unidad asimétrica de un cristal o si, por lo contrario, están limitadas a una cadena. El propio diseño experimental no lo permite debido a los criterios utilizados para la inclusión en el análisis de cada proteína individualmente y de las clases a las que pertenecen. Cuando hay otras proteínas simultáneamente presentes en la celda unidad, éstas no necesariamente satisfacen – y en general no lo hacen– los requisitos para el

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

tratamiento analítico. Es decir que: o bien no contamos con la cantidad mínima requerida de representaciones cristalográficas depositadas en el PDB o, aun cuando se hallaren, no superan en cantidad suficiente los procedimiento de control. Sin embargo, en algunos casos en los que pudo determinarse, se observó cadenas ultracompactas de proteínas pertenecientes a diferentes clases en la misma unidad asimétrica, lo que sugiere que la compactación, si ocurre, podría afectar a más de una estructura en la celda.

5.1.1. Compactación inducida por temperatura

En el curso de esta investigación, demostramos que el R_G de las proteínas cristalizadas es un parámetro sensible a la temperatura de recolección de los datos de difracción. De este modo, probamos que las estructuras resueltas a temperatura ultrabaja ($\leq 160K$) exhiben un R_G significativamente menor al conjunto de las estructuras cuyos patrones de difracción fueron recolectados a temperaturas moderadas o ambientales ($\geq 240K$) (ver figura 4.2). En la misma figura se aprecia con notable claridad el carácter bimodal de la densidad de distribución en las estructuras que fueron resueltas con datos obtenidos a temperaturas intermedias, en el entorno de 160 K y 240 K. El comportamiento de esta curva de distribución en este intervalo es consistente con una "transición de fase" con coexistencia de los dos estados (el normal y el compacto). Se sabe que los polímeros, en general, y las proteínas, en particular, experimentan una transición conocida como *dynamic glass transition* o "transición vítrea" en el rango de 180 – 220 K [13, 60, 11, 65, 20], lo que nos sugiere, por su simultaneidad, una relación entre este fenómeno transicional, ya profusamente descrito, y la compactación estructural que investigamos en este trabajo.

La "transición vítrea" en proteínas –que no es estrictamente una transición termodinámica– refleja la vitrificación del polímero, vinculada con una reducción de los movimientos colectivos

de segmentos de larga escala, que ni siquiera ocurren por debajo de la temperatura de transición, conservándose sólamente movimientos de escasa amplitud, locales y vibracionales. Este incremento de la rigidez de la proteína lleva aparejada una pérdida completa de la función catalítica en el caso de las enzimas. Algunos estudios demuestran que por debajo de la temperatura de transición hay una disminución del volumen de las estructuras específicamente consideradas [27, 68] además de una menor dispersión de la posición de los átomos, que se registra como un estrechamiento de la distribución promedio de los factores de Debye - Waller centrados en valores más bajos.

Esta disminución del volumen y, en consecuencia, del R_G de las estructuras por debajo de la temperatura de la transición vítrea posiblemente se debe al "congelamiento" del movimiento de las regiones móviles (*loops*), movimientos que sí se observan por encima de los 200 K, de acuerdo con lo que indican Tilton et al. [68] en su trabajo de determinación de las estructuras cristalográficas de Rnasa A a nueve temperaturas distintas. De hecho, en ese trabajo los autores observaron que, conforme se incrementa la temperatura, la proteína se expande ligeramente –0.4 % cada 100 K– debido a sutiles diferencias de empaquetamiento que afectan particularmente a los segmentos expuestos [68].

1. En en nuestro trabajamos hallamos que la disminución del $R_G$por debajo de la temperatura de la "transición vítrea" es un fenómeno general con repercusiones en la estabilidad termodinámica de las moléculas, toda vez que va acompañada de un aumento del número de contactos de van der Waals y un acortamiento de la longitud de los puentes de hidrógeno dentro de las estructuras.

5.1.2. Estados nativos ultracompactos

Si acaso existe un factor experimental excluyente que da lugar a la ultracompactación, éste no se pudo identificar a través del análisis estadístico. Dado que el acondicionamiento del cristal que permite la recolección de datos de difracción de rayos X es muy elaborado y propenso a la histéresis [26, 10], pequeñas variaciones en el procedimiento podrían causar que el efecto criogénico alcance su máximo solo en una pequeña fracción de las estructuras resueltas. En cualquier caso, los resultados proporcionados por el estudio estadístico del PDB hacen posible el diseño de experimentos mejor elaborados con el fin de identificar factores metodológicos que conduzcan o propicien la ultracompactación, si éstos existen.

Sin embargo, es posible que la presencia de proteínas con un R_G tan atípicamente bajo, incluso para el grupo de estructuras determinadas por crío-cristalografía, no obedezca a causas únicas o aislables y se comprenda mejor en el contexto teórico de los subestados conformacionales. Bajo este modelo, las estructuras que conservan el plegado tridimensional nativo pero en un epaquetamiento más denso serían "subestados" o isómeros del estado basal: especies atrapadas en irregularidades del mínimo global de la superficie de energía de Gibbs que corresponde a la proteína nativa. Estas estructuras estarían en mínimos dentro de mínimos, en valles separados por pequeñas barreras de energía observables en la estructura fina que constituyen la rugosidad del fondo de la hipersuperficie. Así, un estado ultracompacto se correspondería con un mínimo dentro del mínimo cuyo aislamiento sería imposible a temperaturas más altas, donde las barreras de energía entre esas depresiones sutiles fueran insuficientes para confinar un subestado conformacional.

Más allá de la causa de la ultracompactación, su existencia provoca una serie de preguntas interesantes. Entre ellas, si la ultracompactación es una forma de estabilización energética. La

cantidad y la longitud de interacciones no covalentes son ampliamente utilizadas como criterio principal para evaluar la energía de una conformación proteica. En este sentido, encontramos que las estructuras ultracompactas tienen en promedio más contactos de van der Waals y forman puentes de hidrógeno internos de menor tamaño, incluso cuando se las compara con estructuras resueltas con datos recolectados a temperaturas criogénicas. En consecuencia, es probable que dichas estructuras involucren una forma de compensación a nivel entálpico de la disminución de entropía devenida de las restricciones en la flexibilidad que supone el sobre-empaquetamiento. La otra interrogante fundamental estriba en la excepcionalidad de los hallazgos. ¿Por qué la crío-cristalografía captura el efecto de ultracompactación sólo en muy contadas ocasiones? Acaso porque un subestado conformacional de súper-empaquetamiento es una rareza estadística demasiado extrema. En ese caso, ¿si relajáramos el criterio de selección cuantitativa de las clases de cadenas de proteínas ($\geqslant$ 20 cadenas representadas) saldrían a la luz más casos? En principio, un criterio más débil en la cantidad de representaciones cristalográficas exigidas para incluir en nuestro análisis a la clase a la que pertenecen permitiría incorporar más clases, pero redundaría en una mayor incertidumbre de los hallazgos, porque las clases contendrían, en promedio, menos cantidad de cadenas. Por otra parte, si la ultracompactación, como todo indica, es un fenómeno de muy baja probabilidad, la incorporación de un mayor número de clases no aportaría necesariamente más casos, toda vez que las nuevas clases incorporadas tienen menos estructuras. No obstante, las respuestas a éstas y otras preguntas que deben considerarse requieren otros tipo de enfoques analíticos: por ejemplo, estudios caso a caso mediante simulaciones de dinámica molecular y el diseño de experimentos cuidadosos para estudiar y, más aún, provocar el efecto de ultracompactación en condiciones controladas y reproducibles.

Bibliografía

[1] Anjum Ansari, Joel Berendzen, Samuel F Bowne, Hans Frauenfelder, IE Iben, Todd B Sauke, Erramilli Shyamsunder, and Robert D Young. Protein states and proteinquakes. *Proceedings of the National Academy of Sciences*, 82(15):5000–5004, 1985.

[2] Gustavo A Arteca. Scaling behavior of some molecular shape descriptors of polymer chains and protein backbones. *Physical Review E*, 49(3):2417, 1994.

[3] Rv H Austin, K-W_ Beeson, L Eisenstein, H Frauenfelder, and IC Gunsalus. Dynamics of ligand binding to myoglobin. *Biochemistry*, 14(24):5355–5373, 1975.

[4] João Alexandre RG Barbosa, Luciano P Silva, Rozeni CL Teles, Gisele F Esteves, Ricardo B Azevedo, Manuel M Ventura, and Sonia M de Freitas. Crystal structure of the bowman-birk inhibitor from vigna unguiculata seeds in complex with β-trypsin at 1.55 å resolution and its structural properties in association with proteinases. *Biophysical Journal*, 92(5):1638–1650, 2007.

[5] William S Bennett and Thomas A Steitz. Glucose-induced conformational change in yeast hexokinase. *Proceedings of the National Academy of Sciences*, 75(10):4848–4852, 1978.

[6] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

[7] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The protein data bank. *Nucleic Acids Res.*, 28:235–42, 2000.

[8] Joseph D Bryngelson, Jose Nelson Onuchic, Nicholas D Socci, and Peter G Wolynes. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Structure, Function, and Bioinformatics*, 21(3):167–195, 1995.

[9] G Caracciolo, G Amiconi, L Bencivenni, G Boumis, R Caminiti, E Finocchiaro, B Maras, C Paolinelli, and A Congiu Castellano. Conformational study of proteins by saxs and edxd: the case of trypsin and trypsinogen. *European Biophysics Journal*, 30(3):163–170, 2001.

[10] Leandro Grille Coronel, Juan P Acierno, and Mario R Ermácora. Ultracompact states of native proteins. *Biophysical Chemistry*, 2017.

[11] Stephen Cusack and Wolfgang Doster. Temperature dependence of the low frequency dynamics of myoglobin. measurement of the vibrational frequency distribution by inelastic neutron scattering. *Biophysical Journal*, 58(1):243–251, 1990.

[12] José García de la Torre, María L Huertas, and Beatriz Carrasco. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophysical Journal*, 78(2):719–730, 2000.

[13] Wolfgang Doster. The protein-solvent glass transition. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1804(1):3–14, 2010.

[14] Christian Dumas and Joël Janin. Conformational changes in arginine kinase upon ligand binding seen by small-angle x-ray scattering. *FEBS Letters*, 153(1):128–130, 1983.

[15] Thomas Earnest, Eric Fauman, Charles S Craik, and Robert Stroud. 1.59 å structure

of trypsin at 120 k: comparison of low temperature and room temperature structures. *Proteins: Structure, Function and Bioinformatics*, 10(3):171–187, 1991.

[16] Leif A Eriksson. *Theoretical Biochemistry: Processes and Properties of Biological Systems*, volume 9. Elsevier, 2001.

[17] Gisele Ferreira Esteves, Rozeni Chagas Lima Teles, Nayara Silva Cavalcante, David Neves, Manuel Mateus Ventura, João Alexandre Ribeiro Gonçalves Barbosa, and Sonia Maria De Freitas. Crystallization, data collection and processing of the chymotrypsin–btci–trypsin ternary complex. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 63(12):1087–1090, 2007.

[18] RF Fischetti, DJ Rodi, DB Gore, and L Makowski. Wide-angle x-ray solution scattering as a probe of ligand-induced conformational changes in proteins. *Chemistry & Biology*, 11(10):1431–1443, 2004.

[19] Paul J Flory. The configuration of real polymer chains. *The Journal of Chemical Physics*, 17(3):303–310, 1949.

[20] Hans Frauenfelder, Shirley S Chan, and Winnie S Chan. *The Physics of Proteins: An Introduction to Biological Physics and Molecular Biophysics*. Biological and Medical Physics, Biomedical Engineering. Springer, New York, NY, 2010.

[21] Hans Frauenfelder and PW Fenimore. Física biológica. *Revista Española de Física*, 17(5), 2011.

[22] Hans Frauenfelder, Hermann Hartmann, Martin Karplus, ID Kuntz Jr, John Kuriyan, Fritz Parak, Gregory A Petsko, Dagmar Ringe, and Robert F Tilton Jr. Thermal expansion of a protein. *Biochemistry*, 26(1):254–261, 1987.

[23] Hans Frauenfelder, Fritz Parak, and Robert D Young. Conformational substates in proteins. *Annual Review of Biophysics and Biophysical Chemistry*, 17(1):451–479, 1988.

[24] Hans Frauenfelder, Gregory A Petsko, and Demetrius Tsernoglou. Temperature-dependent x-ray diffraction as a probe of protein structural dynamics. *Nature*, 280(5723):558, 1979.

[25] Hans Frauenfelder, Stephen G Sligar, and Peter G Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603, 1991.

[26] Bertil Halle. Biomolecular cryocrystallography: structural changes during flash-cooling. *Proceedings of the National Academy of Sciences*, 101(14):4793–4798, 2004.

[27] H Hartmann, F Parak, W Steigemann, GA Petsko, D Ringe Ponzi, and H Frauenfelder. Conformational substates in a protein: structure and dynamics of metmyoglobin at 80 k. *Proceedings of the National Academy of Sciences*, 79(16):4967–4971, 1982.

[28] Lizhong He, Sabine André, Hans-Christian Siebert, Heike Helmholz, Bernd Niemeyer, and Hans-Joachim Gabius. Detection of ligand-and solvent-induced shape alterations of cell-growth-regulatory human lectin galectin-1 in solution by small angle neutron and x-ray scattering. *Biophysical Journal*, 85(1):511–524, 2003.

[29] Liu Hong and Jinzhi Lei. Scaling law for the radius of gyration of proteins and its dependence on hydrophobicity. *Journal of Polymer Science Part B: Polymer Physics*, 47(2):207–214, 2009.

[30] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996.

[31] W. Humprey, A. Dalke, and K. Schulten. Vmd-visual molecular dynamics. *J. Mol. Graph.*, 14:33–38, 1996.

[32] Sarah Hunter, Rolf Apweiler, Teresa K Attwood, Amos Bairoch, Alex Bateman, David Binns, Peer Bork, Ujjwal Das, Louise Daugherty, Lauranne Duquenne, et al. Interpro: the integrative protein signature database. *Nucleic Acids Research*, 37(suppl_1):D211–D215, 2008.

[33] Leo C James, Pietro Roversi, and Dan S Tawfik. Antibody multispecificity mediated by conformational diversity. *Science*, 299(5611):1362–1367, 2003.

[34] Daniel A Keedy, Lillian R Kenner, Matthew Warkentin, Rahel A Woldeyes, Jesse B Hopkins, Michael C Thompson, Aaron S Brewster, Andrew H Van Benschoten, Elizabeth L Baxter, Monarin Uervirojnangkoorn, et al. Mapping the conformational landscape of a dynamic enzyme by multitemperature and xfel crystallography. *eLife*, 4:e07574, 2015.

[35] Chae Un Kim, Mark W Tate, and Sol M Gruner. Protein dynamical transition at 110 k. *Proceedings of the National Academy of Sciences*, 108(52):20897–20901, 2011.

[36] S Anand Kumar, N Sanjeeva Murthy, and Joseph S Krakow. Ligand-induced change in the radius of gyration of camp receptor protein from escherichia coli. *FEBS Letters*, 109(1):121–124, 1980.

[37] Yu Chih Liu, Zhenjun Chen, Scott R Burrows, Anthony W Purcell, James McCluskey, Jamie Rossjohn, and Stephanie Gras. The energetic basis underpinning t-cell receptor recognition of a super-bulged peptide bound to a major histocompatibility complex class i molecule. *Journal of Biological Chemistry*, 287(15):12267–12276, 2012.

[38] M Yu Lobanov, NS Bogatyreva, and OV Galzitskaya. Radius of gyration as an indicator of protein structure compactness. *Molecular Biology*, 42(4):623–628, 2008.

[39] Vassiliy Lubchenko, Peter G Wolynes, and Hans Frauenfelder. Mosaic energy landscapes

of liquids and the control of protein conformational dynamics by glass-forming solvents. *The Journal of Physical Chemistry B*, 109(15):7488–7499, 2005.

[40] Francesco Mallamace, Carmelo Corsaro, Domenico Mallamace, Sebastiano Vasi, Cirino Vasi, Piero Baglioni, Sergey V Buldyrev, Sow-Hsin Chen, and H Eugene Stanley. Energy landscape in protein folding and unfolding. *Proceedings of the National Academy of Sciences*, page 201524864, 2016.

[41] Walter F Mangel, BH Lin, and V Ramakrishnan. Characterization of an extremely large, ligand-induced conformational change in plasminogen. *Science*, 248(4951):69–73, 1990.

[42] Joseph A Marsh and Julie D Forman-Kay. Sequence determinants of compaction in intrinsically disordered proteins. *Biophysical Journal*, 98(10):2383–2390, 2010.

[43] Alex L Mitchell, Teresa K Attwood, Patricia C Babbitt, Matthias Blum, Peer Bork, Alan Bridge, Shoshana D Brown, Hsin-Yu Chang, Sara El-Gebali, Matthew I Fraser, et al. Interpro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, 2018.

[44] Osamu Miyashita, José Nelson Onuchic, and Peter G Wolynes. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proceedings of the National Academy of Sciences*, 100(22):12570–12575, 2003.

[45] Sanzo Miyazawa and Robert L Jernigan. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules*, 18(3):534–552, 1985.

[46] ME Newcomer, BA Lewis, and FA Quiocho. The radius of gyration of l-arabinose-binding

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

protein decreases upon binding of ligand. *Journal of Biological Chemistry*, 256(24):13218–13222, 1981.

[47] GA Olah, RD Mitchell, TR Sosnick, DA Walsh, and J Trewhella. Solution structure of the camp-dependent protein kinase catalytic subunit and its contraction upon binding the protein kinase inhibitor peptide. *Biochemistry*, 32(14):3649–3657, 1993.

[48] GA Olah, S Trakhanov, J Trewhella, and FA Quiocho. Leucine/isoleucine/valine-binding protein contracts upon binding of ligand. *Journal of Biological Chemistry*, 268(22):16241–16247, 1993.

[49] José Nelson Onuchic, Zaida Luthey-Schulten, and Peter G Wolynes. Theory of protein folding: the energy landscape perspective. *Annual Review of Physical Chemistry*, 48(1):545–600, 1997.

[50] F Parak, H Hartmann, KD Aumann, H Reuscher, G Rennekamp, H Bartunik, and W Steigemann. Low temperature x-ray investigation of structural distributions in myoglobin. *European Biophysics Journal*, 15(4):237–249, 1987.

[51] Linus Pauling, Robert B Corey, and Herman R Branson. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences*, 37(4):205–211, 1951.

[52] E Pechkova, NL Bragazzi, and C Nicolini. Protein crystallization by anodic porous alumina (apa) template: The example of hen egg white lysozyme (hewl). *NanoWorld J*, 1(2):46–55, 2015.

[53] Cecília Steinberg Perilo, Márcio Tadeu Pereira, Marcelo Matos Santoro, and Ronaldo Alves Pinto Nagem. Structural binding evidence of the trypanocidal drugs berenil® and

pentacarinate® active principles to a serine protease model. *International Journal of Biological Macromolecules*, 46(5):502–511, 2010.

[54] Clifford A Pickover, David B McKay, Donald M Engelman, and Thomas A Steitz. Substrate binding closes the cleft between the domains of yeast phosphoglycerate kinase. *Journal of Biological Chemistry*, 254(22):11323–11329, 1979.

[55] Ingrid PILZ, Faust MALNIG, Otto KRATKY, and Friedrich HAAR. On the conformation of serine-specific transfer rna. *European Journal of Biochemistry*, 75(1):35–41, 1977.

[56] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[57] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

[58] Jane S Richardson. The anatomy and taxonomy of protein structure. In *Advances in Protein Chemistry*, volume 34, pages 167–339. Elsevier, 1981.

[59] Jane S Richardson. Early ribbon drawings of proteins. *Nature Structural & Molecular Biology*, 7(8):624, 2000.

[60] Dagmar Ringe and Gregory A Petsko. The 'glass transition'in protein dynamics: what it is, why it occurs, and how to exploit it. *Biophysical Chemistry*, 105(2):667–680, 2003.

[61] Valeria A Risso, Juan P Acierno, Stefano Capaldi, Hugo L Monaco, and Mario R Ermácora. X-ray evidence of a native state with increased compactness populated by tryptophan-less b. licheniformis β-lactamase. *Protein Science*, 21(7):964–976, 2012.

[62] Benjamin Sandler, Masahiro Murakami, and Jon Clardy. Atomic structure of the trypsin-

aeruginosin 98-b complex. *Journal of the American Chemical Society*, 120(3):595–596, 1998.

[63] Michael Sands, Michael A Kron, and Richard B Brown. Pentamidine: a review. *Reviews of Infectious Diseases*, 7(5):625–6344, 1985.

[64] BA Seaton, JF Head, DM Engelman, and FM Richards. Calcium-induced increase in the radius of gyration and maximum dimension of calmodulin measured by small-angle x-ray scattering. *Biochemistry*, 24(24):6740–6743, 1985.

[65] Jeremy Smith, Krzysztof Kuczera, and Martin Karplus. Dynamics of myoglobin: comparison of simulation results with neutron scattering spectra. *Proceedings of the National Academy of Sciences*, 87(4):1601–1605, 1990.

[66] MM Teeter, A Yamano, B Stec, and U Mohanty. On the nature of a glassy state of matter in a hydrated protein: relation to protein function. *Proceedings of the National Academy of Sciences*, 98(20):11242–11247, 2001.

[67] Daan Thorn Leeson, Douwe A Wiersma, Klaus Fritsch, and Josef Friedrich. The energy landscape of myoglobin: an optical study, 1997.

[68] Robert F Tilton Jr, John C Dewan, and Gregory A Petsko. Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-a at nine different temperatures from 98 to 320k. *Biochemistry*, 31(9):2469–2481, 1992.

[69] Jill Trewhella, DK Blumenthal, SE Rokop, and PA Seeger. Small-angle scattering studies show distinct conformations of calmodulin in its complexes with two peptides based on the regulatory domain of the catalytic subunit of phosphorylase kinase. *Biochemistry*, 29(40):9316–9324, 1990.

[70] Douglas Tsao and Nikolay V Dokholyan. Macromolecular crowding induces polypeptide compaction and decreases folding cooperativity. *Physical Chemistry Chemical Physics*, 12(14):3491–3500, 2010.

[71] Dennis Vitkup, Dagmar Ringe, Gregory A Petsko, and Martin Karplus. Solvent mobility and the protein'glass' transition. *Nature Structural & Molecular Biology*, 7(1):34–38, 2000.

[72] JOACHIM Walter, WOLFGANG Steigemann, TP Singh, HANS Bartunik, WOLFRAM Bode, and ROBERT Huber. On the disordered activation domain in trypsinogen: chemical labelling and low-temperature crystallography. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*, 38(5):1462–1472, 1982.

[73] G. Wang and Jr. Dunbrack, R. L. Pisces: recent improvements to a pdb sequence culling server. *Nucleic Acids Res.*, 33(Web Server issue):W94–8, 2005.

[74] Bojan Zagrovic, Guha Jayachandran, Ian S Millett, Sebastian Doniach, and Vijay S Pande. How large is an α-helix? studies of the radii of gyration of helical peptides by small-angle x-ray scattering and molecular dynamics. *Journal of Molecular Biology*, 353(2):232–241, 2005.

[75] Michael H Zehfus and George D Rose. Compact units in proteins. *Biochemistry*, 25(19):5759–5765, 1986.

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes