

Agradecimientos

A D. José García de la Torre, director de esta Tesis, por acogerme en su grupo de investigación y abrirme las puertas de este mundo apasionante. Mis más sinceros agradecimientos por su confianza, apoyo, sus consejos y por la formación que he adquirido, que me ha posibilitado la realización de esta Tesis.

A D. Álvaro Ortega Retuerta, co-director, le agradezco las lecturas, correcciones, tanto científicas como literarias de esta Tesis, además del seguimiento y la supervisión continua de la misma. Por transmitirme su conocimiento, sus sugerencias e ideas, además de su amistad y compañía en el día a día.

A D. Francisco Martínez Ortiz, anterior director del Departamento de Química-Física, y a Dña. María del Carmen López Martínez, actual Directora, por las facilidades prestadas en la realización de este trabajo. Mención especial a Carmina, por la cercanía diaria y la gran ayuda recibida durante estos cuatro años, en lo personal y en lo referido a gestión y docencia.

A D. José Ginés Hernández Cifre, cuya ayuda y consejos me han facilitado enormemente la elaboración de esta Tesis. Las conversaciones que hemos mantenido me han sido de gran utilidad científica, siendo una importante fuente de información y referencia, además de ser una gran compañero y amigo. A Ricardo, por el soporte informático que he necesitado. A Ana, por la gran ayuda recibida en la última etapa de la Tesis, siempre dispuesta a ayudarme. Y al resto de compañeros y amigos del grupo de investigación de Macromoléculas: Gustavo, Ramón y Guillermo. Gracias a todos por vuestra ayuda, y por todos los momentos de amistad y compañerismo que hemos pasado juntos.

A mis compañeros de departamento: Bolarín, Antonio, Jose Pedro, Miguel, Cerezo, Marwa, Eduardo, Carmen, Encarni y Fran así como a los profesores Joaquín, Carmiña, Pedro Ruipérez, Manuela y Adolfo por la ayuda recibida durante mis labores docentes. Además, también quisiera agradecer a Úrsula y Pedro, personal de administración y servicios, por su gran profesionalidad y calidad humana.

A Lorena, por su cariño, paciencia y comprensión. Gracias por permanecer a mi lado, por tu infinita paciencia y por todo el apoyo y el ánimo recibido durante estos años. A mi familia, especialmente a mis padres y a mis hermanos, por haber estado siempre ahí, dispuestos a ayudar en lo que hiciera falta. Sin ellos no hubiera podido llegar a este punto.

Al resto de personas que he dejado de nombrar por no extenderme en exceso, y a aquellos que han colaborado de una manera u otra y me han acompañado en la elaboración de esta Tesis, Máster y otros trabajos.

Y por último, deseo expresar mi agradecimiento al *Ministerio de Ciencia e Innovación* del gobierno de España por la concesión de ayudas al proyecto CTQ2006-06381 al cuál fue asignada la beca FPI disfrutada por el autor de esta Tesis. Y a la *Fundación Séneca*, órgano oficial para para la investigación científica de la Comunidad Autónoma de la Región de Murcia, cuya financiación proviene de la distinción como “Grupo de Excelencia Científica” (referencia 04531/GERM/06), que me ha permitido disponer de más medios y facilidades, y la posibilidad de extender mi estancia para la finalización de esta Tesis.

Índice general

I. Introducción	19
I.1. Una perspectiva histórica	19
I.2. Macromoléculas rígidas: Teoría y cálculo	21
I.3. Flexibilidad macromolecular	21
I.4. Estructura y propiedades en disolución	22
I.5. Orden y detalle estructural: nuevos paradigmas	23
II. Generalidades y metodología	25
II.1. Propiedades en disolución	26
II.1.1. Propiedades conformacionales	26
II.1.1.a. Radio de giro y propiedades de dispersión estática de luz	26
II.1.2. Propiedades hidrodinámicas	28
II.1.2.a. Coeficiente de difusión traslacional	28
II.1.2.b. Coeficiente de sedimentación	30
II.1.2.c. Dinámica reorientacional e interna	31
II.1.2.d. Viscosidad intrínseca	36
II.1.3. Radios equivalentes y relaciones de radios	37
II.2. Metodologías para partículas rígidas	39
II.3. Metodologías para partículas flexibles	41
II.3.1. Método Monte Carlo	41
II.3.2. Método de Dinámica Browniana	43
III. Avances metodológicos	45
III.1. Aspectos computacionales	45
III.1.1. Optimización del cálculo hidrodinámico	45
III.1.2. Interfaces gráficas para los programas HYDRO++ e HYDROPRO	55

III.2. Mejoras en el tratamiento de la viscosidad	58
III.3. Propiedades de partículas cilíndricas	65
III.4. Potenciales torsionales para cadenas flexibles	71
III.5. HYDROFIT: de las propiedades a la estructura	78
IV. Macromoléculas vermiformes	95
IV.1. Introducción	95
IV.2. Metodología	96
IV.2.1. Modelo vermiforme	96
IV.2.2. Monte Carlo de cuerpo rígido (RBMC) y cálculo de propiedades de modelos de esferas	97
IV.2.3. Versión discreta del modelo vermiforme. La cadena de Schellman . . .	99
IV.2.4. La cadena vermiforme en el límite flexible	102
IV.2.5. Efecto del volumen excluido	104
IV.3. Simulación Monte Carlo	106
IV.3.1. Generación de cadena	106
IV.3.2. Procedimiento de simulación Monte Carlo	107
IV.3.3. Predicción de las propiedades en disolución	109
IV.4. Aplicaciones	111
IV.4.1. Determinación de los parámetros del modelo vermiforme desde sus datos experimentales	111
IV.4.2. Ejemplos. ADN	115
IV.4.3. Otros polímeros	123
IV.4.4. Comparación con otras metodologías	135
IV.4.5. Resumen y conclusiones	137
V. Proteínas	139
V.1. Introducción	139
V.2. Cálculo hidrodinámico de proteínas rígidas	141
V.2.1. Proteínas de tamaño bajo-medio	143
V.2.2. Proteínas de gran tamaño	150
V.3. Estudio conformacional e hidrodinámico de proteínas desnaturalizadas . . .	154
V.3.1. Modelo de simulación de proteína desnaturalizada	154
V.3.2. Propiedades conformacionales de proteínas desnaturalizadas	156

V.3.3. Propiedades hidrodinámicas de proteínas desnaturalizadas	161
V.4. Simulación de proteínas parcialmente desordenadas	165
V.4.1. Simulación de fragmentos flexibles	165
V.4.2. Simulación de fragmentos estructurados	167
V.4.3. Aplicaciones	175
V.4.3.a. Proteína quimérica 1PGA	175
V.4.3.b. Proteína S4	179
V.4.3.c. Proteína ZipA	181
V.4.3.d. Proteína X del virus Sendai	184
V.4.3.e. Proteína MMP-12	190
V.4.3.f. Proteína BTK	193
V.5. Conclusiones	194
Conclusiones	197
Publicaciones resultantes de la Tesis	201
Referencias	203



Glosario

Símbolos

Griegos

- Δt : intervalo de tiempo.
- η : viscosidad de una disolución.
- η_0 : viscosidad del disolvente.
- $[\eta]$: viscosidad intrínseca.
- ϕ : rotación en torno al enlace C α -N en proteínas.
- ψ : rotación en torno al enlace C-C α en proteínas.
- π : número Pi.
- ρ : densidad de la disolución.
- σ_i : radio del elemento esférico i en un modelo de bolas.
- τ_c : tiempo de relajación rotacional.

Latinos

- a_C : radio equivalente para el covolumen.
- a_G : radio equivalente para el radio de giro.
- a_I : radio equivalente para la viscosidad intrínseca.
- a_L : radio equivalente para la dimensión máxima.
- a_R : radio equivalente para el tiempo de relajación rotacional.
- a_{RH} : radio equivalente para el tiempo de relajación rotacional armónico.
- a_T : radio equivalente traslacional.
- a_{TD} : radio equivalente traslacional obtenido del coeficiente de difusión traslacional.
- a_{TS} : radio equivalente traslacional obtenido del coeficiente de sedimentación.
- a_V : radio equivalente para el volumen.
- c : concentración de soluto.

- \mathbf{D}_{ij} : tensor de difusión.
- D_r : coeficiente de difusión rotacional.
- D_t : coeficiente de difusión traslacional.
- D_t^0 : coeficiente de difusión traslacional en condiciones estándar.
- \mathbf{F}_i : fuerza que actúa sobre el elemento i del modelo.
- f_r : coeficiente de fricción rotacional.
- f_t : coeficiente de fricción traslacional.
- k_B : constante de Boltzmann.
- M : peso molecular.
- N : número de bolas en un modelo de bolas.
- N_A : número de Avogadro.
- $P(h)$: factor de forma.
- $p(r)$: función de distribución de distancias entre todos los elementos (puntos) que componen un sistema.
- \mathbf{r}_i : vector posición de la bola i del modelo.
- r_{ij} : distancia entre elementos i y j en un modelo macromolecular.
- R_g : radio de giro característico de una estructura macromolecular.
- s : radio de giro instantáneo en una macromolécula flexible, coeficiente de sedimentación.
- $s_{20,w}$: coeficiente de sedimentación en condiciones estándar.
- T : temperatura absoluta.
- t : tiempo.
- u : covolumen.
- V : volumen, energía potencial.
- \bar{v} : volumen específico parcial.

Acrónimos

- ADN: ácido desoxirribonucleico.
- ARN: ácido ribonucleico.
- DB: dinámica browniana.
- DL: dinámica de Langevin.
- DM: dinámica molecular.
- IH: interacción hidrodinámica.

- LJ: Lennard-Jones (potencial de).
- MC: Monte Carlo.
- VE: volumen excluido.

Acrónimos. Terminología inglesa

- RB: tratamiento de cuerpo-rígido (Rigid-Body).
- RBBD: tratamiento de cuerpo rígido no-IH con muestreo de dinámica browniana (Rigid-Body Brownian Dynamics).
- RBMC: tratamiento de cuerpo rígido con muestreo de Monte Carlo (Rigid-Body Monte Carlo).
- HS: esferas duras (Hard Spheres).
- AUC: ultracentrifugación analítica (Analytical Ultra Centrifugation).
- SANS: dispersión de neutrones de bajo ángulo (Small-Angle Neutron Scattering).
- SAXS: dispersión de rayos-x de bajo ángulo (Small-Angle X-ray Scattering).
- SLS: dispersión estática de luz (Static Light Scattering).
- DLS: dispersión dinámica de luz (Dynamic Light Scattering).
- SEC: cromatografía de exclusión por tamaño (Size exclusion Chromatography).
- FTIR: espectroscopía de infrarrojo por transformada de Fourier. (Fourier Transform Infrared).
- NMR: resonancia magnética nuclear (Nuclear Magnetic Resonance).
- PFG-RMN: gradientes de campo magnético en experimentos de NMR (Pulsed Field Gradient-NMR).
- VIS: viscosimetría (Viscosimetry).



Capítulo I

Introducción

I.1. Una perspectiva histórica

Las macromoléculas desempeñan sus peculiares funciones en sistemas notablemente complejos, como el medio heterogéneo y aglomerado del citoplasma celular, o los materiales poliméricos. Sin embargo, la caracterización de la estructura y propiedades que una macromolécula individual – problema esencial, y previo para comprender su papel en esos medios complejos – ha de hacerse en un medio que refleje únicamente esas características individuales: la disolución diluida.

De igual manera a como Einstein determinó – al margen de las teorías cuánticas – el tamaño de una molécula de sacarosa a partir de su difusividad y de la viscosidad de sus disoluciones acuosas, el concepto de macromoléculas como moléculas de gran tamaño, e incluso los primeros modelos de estas como cadenas moleculares, emanan de estudios (sedimentación, difusión, viscosidad, etc...) de disoluciones diluidas realizados por científicos de la talla de Svedberg, Perrin, Staudinger, Kraemer o Debye.

Hasta, aproximadamente, 1980, las medidas de propiedades en disolución constituyeron el medio fundamental – en algunos casos, el único disponible – para caracterizar la estructura de macromoléculas biológicas, desde la mera determinación de su peso molecular, hasta la determinación de su estructura tridimensional. Para la determinación estructural se disponía de teorías para modelos sencillos: elipsoides o cilindros, para macromoléculas rígidas, o la cadena vermiforme para las semiflexibles. A partir de las propiedades se podía deducir, por ejemplo, si una proteína rígida era globular o fibrosa, y cuáles eran sus dimensiones aproximadas. Tras los trabajos de Bloomfield, van Holde y Dalton, se dispuso

de la metodología de modelado con esferas (*bead modelling*) para construir modelos con mayor detalle, por ejemplo para proteínas o complejos macromoleculares compuestos por subunidades. Unos años después, Bloomfield y García de la Torre introdujeron notables mejoras en los aspectos teóricos y computacionales de estos modelos.

Sin embargo, hacia la mencionada década de los 80, el desarrollo de técnicas de difracción y de rayos X, resonancia magnética nuclear (NMR) , y microscopías avanzadas, hizo que estas técnicas se considerasen como las fuentes básicas de instrumentación estructural. Además, por entonces la instrumentación para medir propiedades en disolución – por ejemplo la ultracentrifugación analítica – había quedado de una u otra manera obsoleta. Por todo ello, el estudio de propiedades en disolución quedó relegado a un segundo plano, si bien para bastantes autores siguió teniendo gran importancia, aunque simplemente fuese para validar las estructuras obtenidas de un cristal, fibra o espécimen de microscopía – es decir, para verificar si son realmente las que presentan en medio acuoso.

Pasados pocos años (hacia el año 2000, por dar una fecha), el interés en las propiedades en disolución se recobró plenamente. Contribuyó a ello la mejora de las técnicas instrumentales, por ejemplo, en las nuevas ultracentrífugas analíticas, y más recientemente, los detectores de propiedades hidrodinámicas y de dispersión de luz acoplados a cromatografía y otros medios separativos. También contribuyeron a ello los avances en los métodos de modelado y cálculo hidrodinámico. Concretamente, se consiguió adaptar estos cálculos a las estructuras de nivel atómico obtenidas por cristalografía o NMR, o a las de microscopía electrónica, resultando factible la predicción de propiedades a partir, directamente, de los datos estructurales suministrados por esas técnicas, lo cual como indicamos es imprescindible para confirmar la validez de la estructura en disolución. Durante la última década se han sucedido, además de nuevos avances en instrumentación, otras mejoras en cuanto a las metodologías de predicción, modelización y al cálculo de propiedades en disolución, siendo la presente Tesis una contribución en este sentido.

En el Capítulo II se describen en forma resumida las propiedades que son objeto de este estudio, así como las metodologías de cálculo disponibles para ellas, que han sido no solamente utilizadas, sino también mejoradas en nuestro trabajo.

I.2. Macromoléculas rígidas: Teoría y cálculo

En efecto, la teoría hidrodinámica de modelos de esferas para macromoléculas en disolución, presenta aspectos particularmente difíciles sobre los que ha sido necesario continuar trabajando, y los métodos computacionales han debido adaptarse a los avances en arquitectura de ordenadores. Estas necesidades son las que han motivado el trabajo expuesto en el Capítulo III. Así, en la sección III.1 se presentan avances relativos a la eficiencia y facilidad de los programas de cálculo. En el plano teórico, un problema a tratar se refiere a la viscosidad intrínseca para modelos de esferas. Mientras que para modelos compuestos por un gran número de esferas pequeñas, como los modelos clásicos de ristras de esferas idénticas en forma de largas varillas rígidas, esta propiedad podía ser adecuadamente tratada, el tratamiento de estructuras compuestas por pocas esferas, o bastante desiguales, daba lugar a resultados un tanto deficientes. Sobre este asunto hemos hecho una contribución en la sección III.2. Combinando este avance con métodos existentes, hemos propuesto un tratamiento general para partículas cilíndricas, en todo el rango de relación longitud-diámetro, desde discos finos hasta largas varillas rectas, que se presenta en la sección III.4.

I.3. Flexibilidad macromolecular

Es bastante habitual enfocar el estudio de macromoléculas en disolución, bien como estructuras (cadenas moleculares) muy flexibles, situación típica en los polímeros sintéticos, o bien como estructuras esencialmente rígidas, originadas por un plegamiento muy específico y compacto, que imparte a las macromoléculas biológicas su peculiar estructura y función.

Estos son puntos de vista realmente extremos – aunque realmente útiles como paradigmas – sobre los que construir la ciencia macromolecular. Pero para diversas circunstancias ha sido necesario elaborar conceptos intermedios, uno de los cuales, extremadamente útil, es el modelo vermiforme para cadenas parcialmente flexibles, que permite modelar desde cadenas extremadamente largas y flexibles como los ADN virales, hasta varillas cortas y rígidas, como los oligonucleótidos de ADN doble-helicoidal, así como cadenas semiflexibles de polisacáridos e incluso ciertos polímeros sintéticos. En bastantes macromoléculas biológicas se encuentra una diferente clase de semiflexibilidad; en lugar de tener una cierto grado de flexibilidad uniformemente distribuida a lo largo de toda la macromolécula –

como la cadena vermiforme – tales macromoléculas contienen tanto regiones muy rígidas como otras muy flexibles. Recientemente se están desarrollando métodos para la predicción de las propiedades en disolución de tales estructuras macromoleculares, basadas en métodos de simulación de Monte Carlo y de dinámica Browniana, de los que se hace una breve mención en la sección II.3. Con vistas a aplicaciones que habrían de hacerse después – para proteínas desordenadas – nos hemos visto necesitados de algunas ampliaciones de estos métodos, en lo relativo a potenciales de torsión en cadenas flexibles, de los que se trata en la sección III.5.

El mencionado modelo clásico de cadena vermiforme presentaba, aún recientemente, algunos defectos, tales como la inadecuada descripción de partículas cortas y rígidas, o del efecto de volumen excluido en largas cadenas flexibles. Otro objetivo de esta Tesis ha sido estudiar el modelo vermiforme mediante los métodos más avanzados de hidrodinámica, y de simulación de Monte Carlo, lo cual se trata en la primera sección del Capítulo IV.

I.4. Estructura y propiedades en disolución

Ya hemos indicado cómo las propiedades en disolución tendrían como objetivo primario obtener información estructural, siendo esto posible en cuanto a aspectos estructurales globales. Pero si se considera la estructura a nivel mucho más detallado, estas propiedades no contienen información suficiente para deducir la estructura a partir de ellas. No obstante, sí permiten elucidar aspectos más detallados que los globales, pero menos que los atómicos, a los cuales estas propiedades son sensibles. Se pueden averiguar o refinar detalles estructurales mediante métodos de tanteo, que persigan ajustar las predicciones resultantes de los cálculos a datos experimentales. Habiendo ya desarrollado ampliamente procedimientos para el cálculo, nos hemos ocupado de elaborar otras herramientas computacionales que ayudan en esos procesos de tanteo y ajuste. Estos nuevos desarrollos, que ya había sido llevado a cabo en nuestro grupo para los modelos sencillos (elipsoidales, vermiformes), los hemos extendido en esta Tesis a los modelos de esferas más generales, construyendo programas analizadores que actúan en conexión con los programas predictores. De ello se trata en la sección III.6, donde se presenta el concepto general y, como ejemplo, la determinación de ciertas características estructurales de anticuerpos. Además, las antes mencionadas mejoras computacionales del modelo vermiforme las hemos conjugado con estas técnicas de análisis, consiguiendo una determinación de los parámetros

estructurales del ADN doble-helicoidal, que ajustan perfectamente datos experimentales de moléculas de ADN desde 8 hasta 2×10^5 pares de base, según se describe en la sección IV.4, conjuntamente con el estudio de otros polímeros.

I.5. Orden y detalle estructural: nuevos paradigmas

Ya hemos indicado como se impuso hacia la década de los 80 un punto de vista según el cual la estructura ordenada, rígida, perfectamente definida de las biomacromoléculas, y su detalle estructural atómico, eran los aspectos fundamentales para entender su función. Desde el punto de vista dinámico, y en particular de la simulación computacional, el mismo paradigma promovió el empleo de métodos de dinámica molecular atomística, posibilitados por el incesante aumento de la potencia de los ordenadores.

Pero se puede afirmar que en la actualidad otros paradigmas están coexistiendo, y en cierto modo, reemplazando al anterior. En cuanto al nivel de detalle, en el ámbito computacional pronto se comprendió que, por mucho avance que tuviera lugar en la potencia de los ordenadores, alcanzar la escala de tiempo de los microsegundos o incluso milisegundos, en los que suceden importantes eventos fisiológicos, era incompatible con conservar el detalle atómico en los modelos. Además se comprendió que el nivel atómico era innecesario para la descripción de ciertos fenómenos, relacionados con aspectos estructurales o dinámicos más amplios. Así, como en otros ámbitos científicos, hay una tendencia actual a emplear modelos de “grano grueso” (*“coarse-grained”*), en los que los elementos representan unidades no tan pequeñas; por ejemplo no átomos, sino residuos completos: aminoácidos o nucleótidos en las biomacromoléculas, o una o unas pocas unidades repetitivas en polímeros. En esta línea, hemos visto conveniente adaptar a esta escala los métodos que anteriormente habíamos pensado para predicción de propiedades a nivel atómico. Así, hemos abordado el cálculo de propiedades en disolución de proteínas y oligonucleótidos a partir de modelos en los que cada elemento representa un residuo, extendiendo de esta manera la metodología a estructuras con menor resolución. De ello se trata en la primera parte del Capítulo V.

En cuanto a la visión de las proteínas como estructuras rígidas, ordenadas, de forma bien definida, en lo cual radica su función, en las últimas décadas se ha puesto de manifiesto la existencia de numerosas proteínas en cuyo estado nativo, funcional, la cadena polipeptídica, en buena parte de su extensión está desplegada. Son las llamadas proteínas

“intrínsecamente desordenadas”. Por ser en menor o mayor grado desordenadas y flexibles, no pueden ser caracterizadas por cristalografía, y las propiedades en disolución, además de manifestar ese grado de desorden, son las herramientas primarias para su caracterización. El poseer tanto zonas quasirígidas como otras muy flexibles hace que el desarrollo de métodos para el cálculo de propiedades de estas proteínas sea un reto – de hecho, un problema para el que no había herramientas computacionales sistemáticas – que hemos abordado en la segunda parte del Capítulo V.

En esta última parte de la Tesis diseñamos un esquema computacional para el cálculo de propiedades en disolución de proteínas desordenadas, comenzando con proteínas que son globulares en forma nativa, pero que bajo ciertos efectos desnaturalizantes, se presentan como flexibles y desordenadas, para pasar después al caso de las denominadas proteínas intrínsecamente desordenadas, que lo son, al menos en parte de su extensión, en su forma nativa. Para este propósito hemos hecho, un esfuerzo de generalización en cuanto al desarrollo de metodologías (conjugando nuestros propios programas con otros de dominio público), para demostrar después el muy satisfactorio rendimiento de esta metodología en numerosos casos concretos.

Capítulo II

Generalidades y metodología

En esta sección se definen las propiedades que se van a manejar a lo largo de esta Tesis, así como las metodologías que serán utilizadas y que ya han sido descritas previamente en multitud de trabajos anteriores. Se hace necesaria una introducción a los parámetros que serán frecuentes en capítulos posteriores y este capítulo, por tanto, servirá a modo de guía y de consulta para el seguimiento de cualquier capítulo contenido en esta Tesis.

Se definen en primer lugar las propiedades características de partículas en disolución a las que se hará mención, incluyendo propiedades conformacionales y propiedades hidrodinámicas. Se describen además los conceptos de los radios equivalentes y relaciones de radios, desarrollado en los últimos años para facilitar y sistematizar el trabajo con las propiedades mencionadas anteriormente. Finalmente se muestran una serie de metodologías de simulación que se emplearán para los diferentes cálculos que se proponen en esta Tesis, desde cálculos de propiedades de partículas rígidas basados en la mecánica de fluidos hasta las conocidas para la realización de barridos conformacionales o de evoluciones temporales de partículas flexibles y el cálculo de sus propiedades, como Monte Carlo (MC) o la Dinámica browniana (DB).

II.1. Propiedades en disolución

II.1.1. Propiedades conformacionales

II.1.1.a. Radio de giro y propiedades de dispersión estática de luz

Un indicador del tamaño (medio) de una cadena macromolecular en disolución, independientemente de la topología de esta, es la propiedad conformacional conocida como radio de giro cuadrático medio, $\langle s^2 \rangle$, definida por la expresión [1]:

$$\langle s^2 \rangle = \frac{1}{N} \sum_{i=1}^N \langle s_i^2 \rangle \quad (\text{II.1})$$

donde s_i^2 es el módulo cuadrático del vector que une el centro de masas de la macromolécula con cualquier elemento puntual i de esta en una conformación dada, siendo $\langle s_i^2 \rangle$ el promedio sobre todas las conformaciones posibles. Otra forma equivalente de calcular el radio de giro cuadrático medio, en términos de los vectores de distancia entre los elementos, es el siguiente [2]:

$$\langle s^2 \rangle = (1/2N^2) \sum_{i=1}^N \sum_{j=1}^N \langle r_{ij} \rangle \quad (\text{II.2})$$

donde r_{ij} indica la distancia entre dos elementos puntuales i, j cualesquiera de la macromolécula, y N es el total de elementos. Sin embargo, lo más usual es referirse al radio de giro, R_g , como la raíz cuadrada del radio de giro cuadrático, que en el caso de moléculas rígidas sería simplemente $R_g = \sqrt{\langle s^2 \rangle}$. Para moléculas flexibles, el radio de giro R_g (promedio), se calcula como $R_g = \sqrt{\langle s^2 \rangle}$. Para el cálculo de $\langle s^2 \rangle$, la segunda expresión (ec. II.2) es más conocida, pero se suele emplear la ecuación II.1 por ser más eficiente computacionalmente.

El radio de giro R_g es accesible experimentalmente empleando técnicas de dispersión de radiación electromagnética (luz visible para partículas grandes o rayos X y neutrones para partículas de pequeño tamaño). En la dispersión estática de radiación electromagnética se tiene en cuenta la intensidad dispersada a cada ángulo como un valor medio de todas las fluctuaciones que se producen en esta. Para descontar la dispersión debida al disolvente y la dependencia de la intensidad dispersada con la distancia a la que está el detector, se define la relación Rayleigh, ΔR :

$$\Delta R = \frac{r^2 \Delta I}{I_0} \quad (\text{II.3})$$

siendo r la distancia de la muestra al detector, I_0 la intensidad incidente e I la intensidad medida. El valor de ΔR a lo largo del rango de ángulos de detección θ varía. Lord Rayleigh demostró que la intensidad difundida total I , cuando se utiliza luz incidente no polarizada, viene dada por:

$$I = \frac{2\pi^2 M I_0}{r^2 \lambda^4 N_A} \left(\frac{dn}{dc} \right)^2 (1 + \cos^2 \theta) \quad (\text{II.4})$$

donde dn/dc es la variación del índice de refracción de la disolución y del disolvente, λ es la longitud de onda y M es la masa molecular. Si ciertos términos se agrupan en la llamada constante óptica K :

$$K = \frac{4\pi^2}{N_A \lambda^4} n_1^2 \left(\frac{dn}{dc} \right)^2 \quad (\text{II.5})$$

entonces, la variación angular en una disolución diluida, según Debye [3], viene dada por:

$$\frac{K}{\Delta R} = \frac{1}{M} \frac{1}{P(\theta)} + 2A_2c \quad (\text{II.6})$$

siendo θ el ángulo al cual se sitúa el detector con respecto a la dirección de la radiación incidente, y A_2 el segundo coeficiente del virial. La variación angular está contenida en el factor de forma $P(h)$, siendo $h = (4\pi/\lambda) \sin(\theta/2)$, que para un sistema de N elementos, viene dada por la fórmula de Debye :

$$P(h) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left\langle \frac{\text{sen } hr_{ij}}{hr_{ij}} \right\rangle \quad (\text{II.7})$$

donde r_{ij} es la distancia entre el elemento i y j . El factor $P(h)$ nos permite conocer el radio de giro cuadrático medio (en el caso de macromoléculas flexibles) o simplemente el radio de giro cuadrático de la macromolécula según la relación:

$$P(h) \simeq 1 - \frac{h^2}{3} \langle s^2 \rangle = 1 - \frac{1}{3} h^2 R_g^2 \quad (\text{II.8})$$

obtenida de un desarrollo de Taylor para valores pequeños de hR_g . La aproximación es válida cuando θ es pequeño, y λ es del orden de magnitud de R_g . Otra aproximación análoga, usada preferentemente en dispersión de rayos X o neutrones, es la ecuación de Guinier,

$$\ln P(q) \simeq -h^2 R_g^2 / 3 \quad (\text{II.9})$$

Otra propiedad que podemos obtener de la dispersión estática de luz es la distribución de distancias entre los posibles pares de puntos dentro de la macromolécula que se podría

expresar como una transformada de Fourier de la forma:

$$P(r) = \frac{1}{2\pi^2} \int_0^\infty (hr)P(h) \sin(hr)dh \quad (\text{II.10})$$

La teoría que gobierna los procesos de dispersión de luz, es la misma para la dispersión de rayos X (SAXS) y neutrones (SANS). Sin embargo, dependiendo del sistema de estudio, más concretamente de las dimensiones de las partículas dispersantes, debemos utilizar la técnica adecuada en cada caso, empleando la radiación cuya longitud de onda (λ) sea inferior a las dimensiones de las partículas en la disolución. Mientras que la longitud de onda de la luz visible está en torno a varios cientos de nanómetros, la longitud de onda de los Rayos X varía en el rango de 1-2 Å, y 2-10 Å en el caso de la radiación de neutrones. Por ello se deduce que, de acuerdo al tipo de haz incidente, varía el rango de magnitud del vector de scattering h , estando en torno a 0.01 - 0.5 Å⁻¹ en el caso de SAXS y SANS, y en el rango 0.0002 - 0.003 Å⁻¹ en el caso de medidas de dispersión de luz estática (SLS).

II.1.2. Propiedades hidrodinámicas

II.1.2.a. Coeficiente de difusión traslacional

La difusión macroscópica, definida como el cambio de la concentración del soluto con la posición y a lo largo del tiempo, es una manifestación del movimiento microscópico –el movimiento browniano– de las partículas. La conexión entre ambos aspectos se establece mediante la teoría de Einstein del movimiento browniano [4], cuyo resultado fundamental es la ecuación de Einstein para el recorrido cuadrático medio $\langle d^2 \rangle$ de una partícula en un tiempo t , dado por [5, 6]:

$$\langle d^2 \rangle = 6D_t t. \quad (\text{II.11})$$

Otra de las contribuciones de la teoría de Einstein es la relación entre el coeficiente de difusión y propiedades hidrodinámicas de la partícula, relacionadas con la fricción que esta experimenta en un líquido viscoso, el disolvente. Cuando una partícula de simetría esférica, se mueve en el seno de un fluido con una velocidad lineal \mathbf{u} , la relación entre la fuerza que dicha partícula experimenta, \mathbf{F} , y la velocidad \mathbf{u} se establece mediante el coeficiente de fricción traslacional, f_t , según la expresión:

$$\mathbf{F} = f_t \mathbf{u}. \quad (\text{II.12})$$

Este coeficiente, f_t , es un escalar y para el caso particular de una partícula esférica viene dado por la ley de Stokes

$$f_t = 6\pi\eta_0\sigma, \quad (\text{II.13})$$

donde η_0 es la viscosidad del disolvente puro y σ es el radio de la esfera. Además, el coeficiente de fricción traslacional f_t , que relaciona la velocidad de una partícula en el seno del disolvente con la fuerza que experimenta dicha partícula, nos permite el cálculo del coeficiente de difusión traslacional D_t gracias a la relación demostrada, de nuevo, por Einstein:

$$D_t = kT/f_t \quad (\text{II.14})$$

donde k es la constante de Boltzmann y T la temperatura. Este coeficiente de difusión representa la migración de las macromoléculas en el seno de un disolvente e interviene en las leyes fenomenológicas de Fick.

Para partículas de forma arbitraria, f_t y D_t se han de sustituir por los correspondientes tensores de fricción y de difusión. El tratamiento es matemáticamente más complicado, pero no presenta ninguna novedad conceptual [7, 8, 9, 10].

En el caso de partículas flexibles, se considera como una buena aproximación el coeficiente de difusión obtenido del promedio conformacional sobre los valores calculados de las conformaciones individuales. Otra posible aproximación es simular la molécula mediante dinámica molecular o browniana, y calcular el desplazamiento cuadrático medio de su centro de masas, obteniendo el coeficiente de difusión empleando la ecuación II.11. Sin embargo, la trayectoria deberá ser lo suficientemente larga para que la molécula flexible adopte todas las conformaciones posibles, y por lo tanto, los diferentes desplazamientos instantáneos debidos a su conformación en un instante dado sean promediados. Para ello, utilizamos la función de correlación para el desplazamiento cuadrático medio:

$$C_{trans}(t) = \langle [r_{cm}(t_0) - r_{cm}(t_0 + t)]^2 \rangle_{t_0} \quad (\text{II.15})$$

Esta expresión es una forma matemáticamente más detallada del desplazamiento cuadrático medio $\langle d^2 \rangle$.

Existen varias técnicas experimentales que nos conducen a la obtención del coeficiente de difusión traslacional, pero la más importante y directa es la denominada dispersión dinámica de luz (DLS). Esta técnica se basa en el hecho de que las partículas en disolución

se encuentran en permanente movimiento debido a los impactos de moléculas de disolvente. Si un haz de luz incide en dichas partículas, la intensidad de luz dispersada muestra variaciones temporales. Mediante DLS se calculan funciones relacionadas con estas variaciones, obteniéndose el coeficiente de difusión traslacional. Otra técnica muy empleada para obtener este coeficiente es la resonancia magnética nuclear (NMR), en su variante de pulsos de gradiente de campo (PFG). Además, D_t está íntimamente relacionado con el coeficiente de sedimentación s , y puede ser calculado a partir de este, como se verá más adelante (ec. II.19).

Por regla general se suele representar el valor de D_t a concentración cero (ya que la concentración puede afectar al valor de la medida) y en condiciones estándar de temperatura y disolvente, 20 °C y agua (20, w). La relación para la conversión es la siguiente:

$$D^{\circ}_{20,w} = D_x \frac{293}{273 + T_x} \frac{\eta_x}{\eta_{20,w}} \quad (\text{II.16})$$

donde x representa las condiciones bajo las que se ha realizado la medida. Para el valor a concentración cero se suelen realizar medidas a varias concentraciones y extrapolar el valor a concentración cero, que es el que nosotros asumiremos en nuestro caso.

II.1.2.b. Coeficiente de sedimentación

Una de las propiedades hidrodinámicas más estudiadas es el coeficiente de sedimentación, para lo que se emplea la ultracentrifugación analítica (AUC), más concretamente, la técnica de velocidad de sedimentación. Mediante esta técnica se somete a la disolución a una centrifugación a regímenes muy altos de velocidad angular de manera que se crea una fuerza centrífuga sobre las partículas en disolución que es opuesta a la fuerza de flotación y que hace que las macromoléculas sedimenten, se forme un frente de concentración que se desplaza hacia el fondo de la célula en que está contenida la muestra. El coeficiente de sedimentación de una molécula se puede definir como:

$$s = \frac{u}{\omega^2 r} \quad (\text{II.17})$$

donde u es la velocidad radial que adquiere de la macromolécula, ω es la velocidad angular del rotor y r es la posición radial.

Una descripción sencilla del fenómeno de la velocidad de sedimentación, que se puede encontrar en los libros de texto [11], relaciona el coeficiente de sedimentación con el

coeficiente de fricción traslacional, f_t , según:

$$s = \frac{M(1 - \bar{v}\rho)}{N_A f_t} \quad (\text{II.18})$$

donde M es el peso molecular de la partícula, \bar{v} es el volumen específico parcial, ρ la densidad de la disolución, aproximadamente igual a la del disolvente y N_A el número de Avogadro. A partir de la ecuación de Einstein para D_t , ec. II.14, se obtiene la conocida ecuación de Svedberg, siendo posible relacionar el coeficiente de sedimentación con el coeficiente de difusión traslacional antes descrito:

$$s = \frac{M(1 - \bar{v}\rho)D_t}{RT} \quad (\text{II.19})$$

donde R indica la constante de los gases perfectos. Como en el caso anterior, el valor de s se suele expresar para disolución acuosa y a 20°C, y suelen realizarse medidas a varias concentraciones para expresar el valor teórico de s a concentración cero. La expresión que nos da el valor en las condiciones estándar antes mencionadas (agua, 20 °C) es:

$$s^{\circ}_{20,w} = s^{\circ}_x \frac{(1 - \bar{v}\rho)_{20,w} \eta_x}{(1 - \bar{v}\rho)_x \eta_{20,w}} \quad (\text{II.20})$$

II.1.2.c. Dinámica reorientacional e interna

El movimiento browniano no tiene solamente un aspecto traslacional. Una partícula en el seno de una disolución, además de trasladarse, ejecutará un movimiento rotacional. La descripción de la difusión rotacional en una partícula de forma arbitraria es más complicada que la traslacional, pues se pueden definir y observar diversos coeficientes (componentes del tensor de difusión rotacional) y tiempos de relajación rotacional. La situación es más sencilla para una partícula isométrica, que tiene un único coeficiente de difusión rotacional, D_r , y un único tiempo de relajación, τ . La teoría de difusión rotacional es también debida a Einstein [4], y establece una relación entre estas dos magnitudes,

$$\tau = \frac{1}{6D_r}, \quad (\text{II.21})$$

donde, como para la traslación, hay una relación difusión-fricción,

$$D_r = \frac{k_B T}{f_r}, \quad (\text{II.22})$$

siendo f_r el coeficiente de fricción rotacional. Para una esfera de radio σ , se tiene que $f_r = 8\pi\eta_s\sigma^3$, y por tanto,

$$\tau = \frac{8\pi\eta_0\sigma^3}{6k_B T} = \frac{V\eta_0}{k_B T}. \quad (\text{II.23})$$

En el caso de partículas rígidas no esféricas, la difusión rotacional está gobernada por el tensor de difusión rotacional \mathbf{D}_{rr} . Este puede obtenerse de fórmulas exactas en ciertos casos (partículas elipsoidales), y en el caso de partículas con forma arbitraria, puede obtenerse mediante el modelado de esferas [12, 13]. \mathbf{D}_{rr} puede ser expresado en términos de sus tres valores propios D_1 , D_2 y D_3 , y sus correspondientes vectores propios, que toman las tres principales direcciones del tensor. El coeficiente de difusión rotacional es la media aritmética de los tres valores propios de \mathbf{D}_{rr} :

$$D_r = \frac{D_1 + D_2 + D_3}{3} \quad (\text{II.24})$$

Los tiempos de relajación pueden calcularse mediante las siguientes expresiones

$$\tau_1 = (6D_r - 2\Delta)^{-1} \quad (\text{II.25})$$

$$\tau_2 = (3(D_r + D_1))^{-1} \quad (\text{II.26})$$

$$\tau_3 = (3(D_r + D_2))^{-1} \quad (\text{II.27})$$

$$\tau_4 = (3(D_r + D_3))^{-1} \quad (\text{II.28})$$

$$\tau_5 = (6D_r + 2\Delta)^{-1} \quad (\text{II.29})$$

siendo Δ la anisotropía de difusividad rotacional de la partícula, que se calcula con la expresión

$$\Delta = \sqrt{D_1^2 + D_2^2 + D_3^2 - D_1D_2 - D_1D_3 - D_2D_3} \quad (\text{II.30})$$

En el caso de partículas flexibles, no se puede hablar de rotación en sentido estricto. La flexibilidad, desde el punto de vista dinámico, consiste en el continuo cambio de

conformación de la molécula, a lo que nos referimos como dinámica interna. Podría calcularse como el promedio conformacional sobre los valores calculados de las conformaciones individuales, tal y como vimos en el cálculo de D_t . Sin embargo, este tratamiento no es apropiado en este caso, salvo que se trate de partículas con un bajo grado de flexibilidad [14].

Sin embargo, mediante simulación molecular o browniana, sí es posible estudiar una dinámica de rotación —que denominamos más específicamente dinámica reorientacional— de un elemento, segmento, porción, etc...— de la molécula o de su modelo físico (por ejemplo, un enlace químico entre dos átomos consecutivos). La dinámica teórica, simulada o experimental suele expresarse en la forma de funciones de correlación, cuya forma general puede expresarse como

$$C(t) = \langle F(t_0, t_0 + t) \rangle_{t_0}, \quad (\text{II.31})$$

donde $F(t_0, t_0 + t)$ es cierta cantidad dependiente de las conformaciones de la molécula en un instante t_0 y en otro instante posterior, $t_0 + t$, promediándose sobre todas las posibles elecciones del instante inicial t_0 , resultando así $C(t)$ una función del tiempo transcurrido, t . La evolución de la orientación de un vector cualquiera \mathbf{v} tiene el típico aspecto de fluctuación aleatoria. Pero en esa fluctuación hay información relevante que se puede extraer mediante funciones de correlación. Por ejemplo, podemos definir la función

$$C_{1,v}(t) = \langle \mathbf{v}(t_0) \cdot \mathbf{v}(t_0 + t) \rangle_{t_0}, \quad (\text{II.32})$$

que decrece monótonamente con el tiempo desde el valor v^2 para $t = 0$ a cero para $t \rightarrow \infty$, o normalizando respecto a v^2 ,

$$P_{1,v}(t) \equiv \frac{C_{1,v}(t)}{v^2} = \langle \cos \beta(t_0, t_0 + t) \rangle_{t_0}, \quad (\text{II.33})$$

siendo $\cos \beta(t_0, t_0 + t)$ el coseno del ángulo formado por dos orientaciones sucesivas del vector, esto es, $\cos \alpha(t_0, t_0 + t) = \mathbf{v}(t_0) \cdot \mathbf{v}(t_0 + t) / v^2 = \mathbf{u}(t_0) \cdot \mathbf{u}(t_0 + t)$, donde $\mathbf{u} = \mathbf{v}/v$ es el vector unitario en la dirección de \mathbf{v} . Otra función de correlación muy utilizada para caracterizar el movimiento browniano reorientacional es el segundo polinomio de Legendre en dicha variable angular,

$$P_{2,v}(t) = \frac{3\langle \cos^2 \beta(t_0, t_0 + t) \rangle_{t_0} - 1}{2}. \quad (\text{II.34})$$

Las funciones de correlación pueden, empíricamente, ajustarse a funciones multiexponenciales, obteniendo los tiempos de relajación del ajuste. Por ejemplo, para partículas

esféricas, la función $P_{2,v}(t)$ (lo mismo es válido para $P_{1,v}(t)$) para cualquier vector radial puede ajustarse a:

$$P_2(t) = e^{(-t/\tau_{esf})} \quad (\text{II.35})$$

y obtener τ_{esf} a partir de un ajuste monoexponencial de la función de correlación, que se define como el tiempo en que la molécula tiene una rotación cuadrática media de un radián. En el caso de partículas rígidas de forma arbitraria (anisométricas), podemos obtener hasta cinco tiempos de relajación:

$$P_2(t) = \sum_{i=1}^5 a_i e^{(-t/\tau_i)} \quad (\text{II.36})$$

que son los formulados en las ecuaciones II.25-II.29, siendo a_i en la ec. II.36 la amplitud asociada a cada tiempo, que depende de los componentes del vector unitario a lo largo del enlace que está siendo monitorizado, cumpliéndose la condición $\sum_{i=1}^5 a_i = 1$. Los tiempos de relajación se clasifican como $\tau_1 \geq \tau_2 \geq \tau_3 \dots$, de manera que τ_1 es el tiempo de relajación más largo. Si la partícula no es esférica, también existe la posibilidad de expresar la dinámica rotacional en términos de un tiempo de correlación efectivo τ_c (o τ_h):

$$\frac{1}{\tau_c} = \frac{1}{5} \sum_{i=1}^5 \frac{1}{\tau_i} = 6D_r \quad (\text{II.37})$$

o bien obtener τ_c a partir de la función de correlación:

$$\tau_c = \int_0^{\infty} P_2(t) dt \quad (\text{II.38})$$

La difusión rotacional puede ser observada experimentalmente a través de la resonancia magnética nuclear (NMR) dinámica, típicamente en partículas rígidas de forma arbitraria (proteínas y fragmentos cortos de ADN). En el caso de proteínas, frecuentemente, se observa la interacción dipolar entre un protón (^1H) y otro núcleo como ^{13}C o ^{15}N , siendo ^{15}N el más usual. Los procesos de relajación spin-red y spin-spin (T_1 y T_2 respectivamente), así como el efecto Overhauser nuclear (NOE) son calculados a partir de la función espectral $J(\omega)$ [15], relacionada con la dinámica rotacional mediante esta relación fundamental

$$J(\omega) = \int_0^{\infty} \langle P_2[\mu(0)\mu(t)] \rangle \cos(\omega t) dt \quad (\text{II.39})$$

siendo μ un vector que define la orientación en el eje de la interacción de relajación (por ejemplo, el vector $^1\text{H}-^{15}\text{N}$). Como hemos visto anteriormente, la función P_2 puede ajustarse a funciones multiexponenciales (ec. II.36), y obtener los tiempos de relajación y sus amplitudes. Para una partícula rígida y anisométrica, $J_0(\omega)$ es una suma de cinco lorentzianas [13]:

$$J_0(\omega) = \sum_{i=1}^5 \frac{a_i \tau_i}{1 + \tau_i^2 \omega^2} \quad (\text{II.40})$$

Si empleamos el tiempo de correlación o armónico (τ_c , ec. II.37), la función de densidad espectral quedaría así :

$$J_0(\omega) = \frac{\tau_c}{1 + \tau_c^2 \omega^2} \quad (\text{II.41})$$

Las magnitudes T_1 , T_2 y NOE estan relacionadas con dicha función de densidad espectral $J(\omega)$, que es una combinación lineal de los valores $J(\omega_H - \omega_X)$, $J(\omega_H)$, $J(\omega_X)$, $J(\omega_H + \omega_X)$ y $J(0)$, siendo ω_H y ω_N la frecuencia característica de Larmor del protón y del núcleo ^{15}N respectivamente:

$$\frac{1}{T_1} = d^2 [J(\omega_N - \omega_H) + 3J(\omega_N) + 6J(\omega_N + \omega_H)] + c^2 J(\omega_N) \quad (\text{II.42})$$

$$\frac{1}{T_2} = (1/2)d^2 [4J(0) + J(\omega_N - \omega_H) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_N + \omega_H)] \quad (\text{II.43})$$

$$NOE = 1 + (\gamma_H/\gamma_N)d^2 [6J(\omega_N + \omega_H) - J(\omega_N - \omega_H)]T_1 \quad (\text{II.44})$$

donde

$$d^2 = (1/10)(\mu_0/4\pi)\hbar^2\gamma_H^2\gamma_N^2r_{NH}^{-6} \quad (\text{II.45})$$

$$c^2 = (2/15)\omega_N^2(\Delta\sigma)^2 \quad (\text{II.46})$$

siendo $\hbar = h/2\pi$ (h =constante de Planck), γ_N y γ_H los radios giromagnéticos de ^{15}N y ^1H respectivamente, r_{NH} la longitud del enlace N-H y $\Delta\sigma$ es la anisotropía del tensor de desplazamiento químico del spin ^{15}N .

Si el enlace N-H perteneciera a una estructura absolutamente rígida, $J(\omega)$ se obtendría directamente de las ecuaciones II.40 o II.41. Sin embargo, en la mayoría de los casos prácticos, el enlace N-H presenta un movimiento interno (del orden de picosegundos) añadido a la difusión rotacional global que experimenta la molécula (del orden de nanosegundos), que hemos definido como τ_c . Por lo tanto, el tiempo de relajación observado para el vector N-H (τ) sería una combinación de τ_c y el debido a la dinámica interna del enlace N-H, que denominamos τ_e :

$$\frac{1}{\tau} = \frac{1}{\tau_c} + \frac{1}{\tau_e} \quad (\text{II.47})$$

y respecto a la función de correlación, según la teoría de Lipari y Szabo [16, 17]:

$$P_2(t) = S^2 e^{(-t/\tau_c)} + (1 - S^2) e^{(-t/\tau)} \quad (\text{II.48})$$

siendo S^2 el denominado *parámetro de orden generalizado*, que depende de la libertad de movimiento que presente dicho enlace, siendo $S^2 = 1$ cuando el enlace es completamente rígido, reduciéndose su valor hasta aproximarse a cero a medida que el enlace es más flexible. En proteínas globulares, el valor de S^2 generalmente oscila entre 0.8-1 en el caso del enlace N-H del esqueleto peptídico, excepto en zonas de gran movilidad como bucles (*loops*) y residuos terminales. En otras proteínas, cuya movilidad interna es elevada o bien presentan regiones desordenadas, es frecuente encontrar residuos con S^2 igual o menores de 0.5 [18]. Sin embargo, es posible eliminar el efecto del movimiento interno mediante el ratio $T1/T2$:

$$T1/T2 = [4J_0(0) + J_0(\omega_X - \omega_H) + 3J_0(\omega_X) + 6J_0(\omega_H) + 6J_0(\omega_X + \omega_H) + (c^2/d^2)((4J_0(0) + 3J_0(\omega_X)))] / [2J_0(\omega_X - \omega_H) + 6J_0(\omega_X)] \quad (\text{II.49})$$

En conclusión, los tiempos de relajación τ_x son una propiedad hidrodinámica adicional, que puede resultar muy útil para caracterizar de manera global la dinámica interna de la molécula.

II.1.2.d. Viscosidad intrínseca

La viscosidad, η , es una medida de la resistencia que oponen a fluir en régimen laminar unas capas de un fluido sobre otras. Es una propiedad que en el caso de disoluciones

macromoleculares conlleva un aumento apreciable con respecto al disolvente puro debido a interacciones de las macromoléculas con el disolvente que hacen que aumenten las fuerzas de rozamiento. La propiedad que caracteriza a la macromolécula en la disolución es la viscosidad intrínseca, $[\eta]$, que se define como:

$$[\eta] = \lim_{c \rightarrow 0} \frac{\eta_{sp}}{c} = \lim_{c \rightarrow 0} \frac{1}{c} \left(\frac{\eta - \eta_0}{\eta_0} \right) \quad (\text{II.50})$$

donde η_{sp} es la viscosidad específica, es decir $\eta_{sp} = \frac{\eta - \eta_0}{\eta_0}$, η_0 es la viscosidad del disolvente puro y c es la concentración. Al hacer el límite a concentración cero se eliminan las interacciones entre moléculas de soluto, lo que hace que $[\eta]$ represente la contribución individual de la molécula a la viscosidad de la disolución. A diferencia de D_t y s , $[\eta]$ depende escasamente de la temperatura y del disolvente (sobre la base de que éstos no afectan notablemente la estructura global de la macromolécula).

Para la medida de viscosidad de disoluciones se emplea tradicionalmente un viscosímetro capilar de vidrio, aunque recientemente se han presentado instrumentos más sofisticados. Conocida la viscosidad de la disolución, la obtención de la viscosidad intrínseca es sencilla, si se realizan medidas a varias concentraciones de la disolución macromolecular (la viscosidad varía con la concentración), es posible, empleando ecuaciones descritas en la bibliografía [19, 20] y que establecen la dependencia de la viscosidad con la concentración, obtener el valor de $[\eta]$.

II.1.3. Radios equivalentes y relaciones de radios

Los radios equivalentes [21] son una forma sistemática de expresar las propiedades en disolución. Se define *radio equivalente* como el radio de la esfera (sin que la partícula tenga en absoluto por qué ser una esfera) que tendría un valor de la propiedad igual a la partícula en estudio. Tales radios se representan como a_X , siendo X el símbolo que se emplea para cada propiedad. Cada una de las propiedades es tomada en la forma de radio equivalente, a_X . Pueden definirse radios equivalentes para cada propiedad, fricción traslacional, a_T , viscosidad intrínseca, a_I , radio de giro a_G , etc. En la Tabla II.1 se muestran los radios equivalentes para algunas propiedades. Estos radios dependen del tamaño y forma de la partícula, de una manera más uniforme que las propiedades mismas. Además, todos los valores de los radios equivalentes, obtenidos a partir de propiedades físicas diferentes, se expresan en las mismas unidades de longitud, lo cual da coherencia a esta manera de representar diversas propiedades.

II.1. PROPIEDADES EN DISOLUCIÓN

Propiedad	Símbolo	Acrónimo	Radio equivalente
Coefficiente de fricción traslacional	f_t	T	$a_T = f_t/6\pi\eta_0$
Tiempos de relajación rotacional	τ	R	$a_R = kT(3/4\pi)^{1/3}(\tau/\eta_0)^{1/3}$
Viscosidad intrínseca	$[\eta]$	I	$a_I = (3/10\pi N_A)^{1/3}([\eta]M)^{1/3}$
Covolumen	u	C	$a_C = (3/32\pi)^{1/3}u^{1/3}$
Volumen	V	V	$a_V = (3/4\pi)^{1/3}V^{1/3}$
Radio de giro	R_g	G	$a_G = \sqrt{5/3}R_g$
Distancia más larga	L	L	$a_L = L/2$

Tabla II.1: Expresiones para los radios equivalentes en varias propiedades en disolución.

Cada par de radios puede combinarse en *relaciones de radios*, $XY = a_X/a_Y$. Los valores a_X , para las diversas propiedades X , no son iguales, pero tampoco difieren mucho, por lo que los valores XY tendrán un valor próximo a la unidad, adimensionales, y no dependerán del tamaño, sino solamente de la forma del soluto macromolecular, es decir, el valor de la expresión no variará con una expansión uniforme e isomorfa de la partícula.

Para una partícula esférica, todos los radios equivalentes son evidentemente iguales al radio geométrico de la propia esfera, y todas las relaciones de radios iguales a la unidad. Para cualquier otra estructura, los valores de los radios equivalentes no son iguales, pero tampoco muy diferentes. Es por ello por lo que los radios equivalentes o relaciones de radios se pueden emplear en lugar de las propiedades en ciertos casos.

II.2. Metodologías para partículas rígidas

En el Grupo de Investigación se vienen desarrollando una serie de programas que emplean la metodología del modelo de esferas (“bead model”) para hacer sencillo y rápido el cálculo de propiedades rígidas de topología arbitraria en disolución. Estos programas escritos en código Fortran están a disposición pública para varias plataformas computacionales (Windows, Linux) y son útiles para el cálculo de propiedades de proteínas, ácidos nucleicos y cualquier otra macromolécula o polímero en disolución, de la cual se conozca su estructura, que debe ser necesariamente rígida.

- HYDRO++ [22, 23] Este programa emplea la metodología del modelo de bolas en sentido estricto, permite el cálculo de un gran número de propiedades: todas las comentadas en el primer apartado de este capítulo, incluidas las derivadas de las medidas mediante dispersión estática de luz. El programa necesita un fichero de entrada con las características del soluto y del disolvente, y un fichero estructural en el cual se especificarán las coordenadas de las esferas que componen el modelo y sus radios. Como resultado se obtiene un fichero de salida que contiene todas las propiedades calculadas y un fichero que permite la visualización del modelo diseñado.
- HYDROPRO [12] En los últimos tiempos se ha logrado obtener mediante diversas técnicas como cristalografía de rayos X o por espectroscopia de NMR (y ha sido publicada) la estructura atómica de gran cantidad de proteínas, agregados supra-moleculares, ácidos nucleicos y muchos más ejemplos de biopolímeros. Esta información puede ser empleada para el modelado de la macromolécula con objeto de su simulación para el cálculo de propiedades. La metodología del modelo de concha (variante del modelo de esferas que modela la superficie de la partícula con pequeñas esferas) para macromoléculas rígidas se puede emplear de manera sencilla utilizando el programa HYDROPRO, que permite basarse en la estructura atómica de la macromolécula. El procedimiento que este sigue es el siguiente: En primer lugar es necesario obtener el modelo atómico, que deberá ser un fichero en el formato de la *Protein Data Bank* (PDB) [24], que es una base de datos pública de proteínas y macromoléculas biológicas. Alternativamente, el usuario puede construir su propia estructura atómica, y crear su propio fichero PDB. El programa coloca una esfera en lugar de cada átomo que no sea de hidrógeno, cuyo radio, a , incluirá la hidratación

que tenga la molécula en disolución. Las esferas solaparán, pues el radio de estas es mayor al radio de van der Waals típico de los átomos de las biomoléculas. En un segundo paso se substituye el volumen que supone este modelo atómico primario por una concha de pequeñas esferas y se obtienen las propiedades calculadas. Por tanto es necesario un fichero con las coordenadas atómicas más el fichero de datos igual que en el programa HYDRO, los resultados se expresan igual que en este programa. Tomando a como un parámetro ajustable, se había encontrado que los valores de este radio atómico hidrodinámico están en un rango relativamente estrecho, centrado en torno a $a = 3.2 \text{ \AA}$, que se consideraba como el radio típico a emplear para hacer predicciones con HYDROPRO. Según indicaremos, el modo de trabajo de este programa y el parámetro a se han refinado en esta Tesis.

- HYDROSUB [25] Aunque es posible la modelización utilizando estructuras de alta resolución como hemos visto con HYDROPRO, hay ocasiones en que es imposible acceder a la estructura atómica de la partícula, en los casos en que las proteínas presentan zonas menos rígidas que hacen imposible su cristalización o moléculas con pesos moleculares muy elevados, fuera del rango de las técnicas cristalográficas o NMR. Por ello puede ser útil un programa que nos permita modelar estructuras complicadas utilizando composiciones de modelos geométricos sencillos, aunque será imposible conseguir una alta resolución. HYDROSUB nos permite utilizar elipsoides tanto prolatos como oblatos (y por tanto esferas) y cilindros para construir nuestro modelo fácilmente, el modelo estará especificado en un fichero estructural, en el que constarán el número de subunidades, su forma y dimensiones y la orientación relativa en el espacio. El programa creará un modelo de concha de la estructura formada por estas subunidades y calculará según este modelo las propiedades en disolución. En una versión posterior [26] se incluye la posibilidad de introducir esferas en el modelo que se suman al cálculo del modelo de concha de las subunidades geométricas con un cálculo en sentido estricto, de esta manera podemos simular una estructura utilizando subunidades elipsoidales o cilíndricas y esferas sólidas y el modelo será un híbrido del modelo de concha con el modelo de esferas en sentido estricto.

II.3. Metodologías para partículas flexibles

Existe la posibilidad de representar una estructura flexible mediante un modelo de esferas, al que se añaden conectores, de flexibilidad variable, y una diversidad de interacciones intramoleculares, representadas por los correspondientes potenciales o fuerzas.

II.3.1. Método Monte Carlo

El método Monte Carlo (MC) es actualmente una de las herramientas más potentes de simulación para predecir propiedades de sistemas físicos complejos, en especial, de macromoléculas flexibles en disolución. Su objetivo en nuestro caso particular de partículas flexibles es simular, a partir de un generador de números aleatorios, una serie de conformaciones del modelo, con el fin de realizar una estadística de las propiedades de interés una vez concluida dicha generación. Existen diversas variantes del método Monte Carlo que emplearemos a lo largo de esta Tesis. En el primero de ellos, el método Monte Carlo más básico, cada conformación generada será completamente independiente de las demás. El promedio, realizado *a posteriori* de los valores correspondientes a cada conformación tendrán como peso estadístico la exponencial de Boltzmann de energía potencial. Las conformaciones totalmente prohibidas son rechazadas.

Uno de estos métodos que emplearemos en esta Tesis es el que denominaremos método Monte Carlo-CARP. Este método de generación de cadenas de rotación libre está basado en la subrutina de cálculo ATCOOR [27]. Conocida la posición de tres elementos consecutivos en la cadena (i , $i + 1$ e $i + 2$), calcula la posición, en coordenadas cartesianas, del cuarto elemento $i + 3$, a una distancia b del elemento $i + 2$, con un ángulo α entre los elementos $i + 1$, $i + 2$ e $i + 3$, y un ángulo de torsión ϕ entre los 4 elementos. Los valores de b , α y ϕ pueden ser fijos, o variables, siguiendo una distribución aleatoria de cualquier tipo. De nuevo, las conformaciones totalmente prohibidas (por ejemplo, por solapamiento) pueden ser rechazadas.

Sin embargo, consideramos que el tratamiento de Monte Carlo más específico y más próximo a la descripción real de una macromolécula flexible es el conocido como método de muestreo por importancia *Metropolis* [28]. En él, la generación de conformaciones se realiza de acuerdo con sus diferentes probabilidades (que se tienen en cuenta *a priori*, en la generación) y es un procedimiento secuencial, a partir de cada conformación aceptada se generará la siguiente. Es imprescindible calcular la energía potencial de cada conforma-

ción generada, lo que nos obliga a establecer las diferentes contribuciones a esta energía, principalmente por las interacciones intramoleculares. Una vez que es generada una conformación, se obtiene la energía potencial, que llamaremos V_{prev} , a continuación, la nueva conformación se obtiene añadiendo a cada coordenada de cada elemento en la conformación anterior un pequeño desplazamiento, δ , que es un número aleatorio uniforme en el intervalo $(-\Delta, \Delta)$, siendo Δ el llamado tamaño de paso de Monte Carlo. Una vez generada una nueva conformación, tendrá asociada una energía potencial, V_{nuev} . Aquí es donde el muestreo por importancia tiene su función, se comparan los dos potenciales, si el nuevo es inferior $V_{nuev} < V_{prev}$, la conformación recién generada se acepta y se almacena para su posterior empleo en el cálculo de propiedades; en caso contrario se establece un examen de aceptación: se genera un número aleatorio uniforme, u , en el intervalo $[0, 1]$; si $u < \exp[-(V_{nuev} - V_{prev})/k_B T]$, la conformación se acepta, en el caso contrario, se rechaza y se retorna a la conformación anterior (es decir $V_{nuev} = V_{prev}$). Mediante este método, cada conformación generada será muy similar a la anterior, por lo que debemos emplear un número relativamente grande de pasos para garantizar la variabilidad conformacional, y elegir una muestra menor, formada por conformaciones separadas entre sí por un elevado número de pasos de Monte Carlo, para así garantizar que esas conformaciones sean prácticamente independientes. Estas conformaciones obtenidas constituyen un conjunto aceptable y bastante acorde con la estadística de Boltzmann asociada a fuerzas intramoleculares.

Independientemente de que se obtengan las conformaciones por un método u otro, se calculan las propiedades de la macromolécula flexible; las propiedades conformacionales, y algunas propiedades hidrodinámicas globales (a las que este método, que no es realmente dinámico, se pueden aplicar), concretamente el coeficiente de difusión traslacional y la viscosidad intrínseca se pueden evaluar —si no exactamente, sí al menos con mucha precisión— también como promedios conformacionales sobre los valores calculados para cada conformación como si fuese rígida, empleando para el cálculo hidrodinámico los tratamientos indicados en la sección II.2, y alguno de los programas de la serie HYDRO [22], tal como el HYDRO++ mencionado también la sección anterior. A este procedimiento de cálculo de propiedades lo denominaremos RBMC, pues combina el método de Monte Carlo con el tratamiento de cuerpo rígido (RB: “rigid body”).

Programa MONTEHYDRO

Para facilitar la tarea de generación de conformaciones y cálculo de propiedades, existe un programa desarrollado en el Grupo de investigación, MONTEHYDRO [29] que implementa una simulación de Monte Carlo del tipo *Metropolis* (muestreo por importancia), sobre un modelo de esferas y conectores. MONTEHYDRO trabaja con las coordenadas cartesianas de las esferas que componen el modelo, modificándolas secuencialmente. Como hemos explicado, tras cada generación, se efectúa la comparación de los potenciales con los de la generación precedente, para ello es imprescindible estimar la energía potencial de cada una, estimación que se hace conociendo la suma de los potenciales de interacción entre las esferas que componen el modelo. A cada tipo de interacción le asignamos un potencial. MONTEHYDRO dispone de una serie de potenciales programados: Potenciales de enlace, de ángulo, de torsión y de elementos no enlazados.

II.3.2. Método de Dinámica Browniana

Aunque el método de MC permite generar posibles conformaciones del sistema físico, no tiene en cuenta la dinámica con la que unas conformaciones se convierten en otras. Por ello, mientras que es válido para estudiar propiedades de equilibrio (o las hidrodinámicas globales con el tratamiento RBMC), no sirve sin embargo para estudiar la dinámica completa —sobre todo la dinámica rotacional o interna— del sistema, salvo que sea el caso de estructuras poco flexibles [14]. Una alternativa al método RBMC es la dinámica browniana (DB), con la que podemos estudiar la evolución temporal de la molécula. Se basa en una ecuación diferencial estocástica que admite una solución algorítmica, que permite simular una trayectoria del modelo, es decir, una serie de posiciones (coordenadas cartesianas) de las esferas, a lo largo del tiempo. A partir de una conformación inicial, el proceso se discretiza en pasos temporales, Δt . A partir del vector de posición de cada esfera en el instante t , \mathbf{r}_i^0 , se calcula el vector posición en $t + \Delta t$ añadiendo ligeros desplazamientos, unos debidos a las fuerzas internas y externas, y otro debido al movimiento browniano propiamente dicho, de acuerdo con una expresión de la forma:

$$\mathbf{r}_i = \mathbf{r}_i^0 + [\Delta \mathbf{r}_i]^{fuerzas} + [\Delta \mathbf{r}_i]^{brow}. \quad (\text{II.51})$$

El algoritmo (de primer orden) inicialmente propuesto para esta finalidad por Ermak y McCammon [30], fue mejorado posteriormente por Iniesta y García de la Torre, que se caracteriza por incluir un paso predictor y otro corrector [31].

Ambos métodos (MC y DB) ignoran la naturaleza molecular del disolvente (considerándolo nuevamente como un fluido continuo de cierta viscosidad), aunque podemos aceptar que este efecto es de segundo orden y de menor importancia que la flexibilidad conformacional del propio soluto. Sin embargo, en un sistema real, las partículas interactúan a través del fluido creando un campo de velocidades que afecta sensiblemente al movimiento del resto de las partículas vecinas, siendo su efecto determinante en la dinámica y en el cálculo de propiedades. Este efecto es denominado interacción hidrodinámica (IH), y se introduce en los algoritmos de DB por medio de tensores de difusión, computacionalmente costosos, calculados mediante interacciones entre pares de partículas. Sin embargo, ignorar la IH hace que la trayectoria no sea representativa de la dinámica real del sistema, aunque sirve para generar conformaciones instantáneas en equilibrio con una estadística conformacional correcta. Así las simulaciones de DB-noIH son válidas para hacer barridos conformacionales, constituyendo alternativas útiles a los métodos de Monte Carlo (de hecho algunos autores las llaman métodos “inteligentes” de Monte Carlo [32]). Sin embargo, la dinámica reorientacional e interna requerirá inevitablemente la inclusión de la IH.

Programa SIMUFLEX

Nuestro Grupo ha desarrollado un paquete de programas, llamado **SIMUFLEX**, para la simulación de trayectorias de DB y el análisis de las mismas. El principal programa, **BROWFLEX**, se encarga de la generación de la trayectoria browniana. Trabaja sobre un modelo mecánico generalizado, por lo que los mismos modelos y parámetros pueden ser utilizados indistintamente en **MONTEHYDRO** (método Monte Carlo) y **BROWFLEX**. El programa **ANAFLEX** se encarga del análisis de las trayectorias, realizando promedios conformacionales (tratamiento RB), funciones de correlación y tiempos de relajación, etc. Las líneas generales de estos programas fueron ya publicadas [33, 34], y recientemente apareció la publicación detallada [35].

Capítulo III

Avances metodológicos

III.1. Aspectos computacionales

III.1.1. Optimización del cálculo hidrodinámico

Pese a los grandes avances en las tecnologías de la computación, el tiempo de cálculo continúa siendo el factor limitante en la gran mayoría de los trabajos teóricos que requieren computación intensiva. El incremento exponencial en la última década de la frecuencia de ciclos del procesador y de la memoria de acceso aleatorio (RAM), unido a la posibilidad de paralelizar ciertas etapas del cálculo, obliga a realizar una revisión exhaustiva de la programación que interviene en el cálculo hidrodinámico, realizada años atrás y evidentemente diseñada para trabajar en máquinas más antiguas que carecían de estos avances. Pretendemos implementar estas mejoras, no sólo en los programas de cálculo a nivel interno del grupo, si no también en los programas que hacemos, de dominio público, para la comunidad científica.

Paralelización del cálculo hidrodinámico

Uno de los avances más importantes en la tecnología de la computación ha sido el desarrollo de la computación paralela. La computación paralela es una técnica de programación en la que muchas instrucciones se ejecutan simultáneamente. Se basa en el principio de que los problemas grandes se pueden dividir en partes más pequeñas que pueden resolverse de forma concurrente ('en paralelo'). La computación paralela emplea elementos de procesamiento múltiple simultáneamente para resolver un problema. Esto se

logra dividiendo el problema en partes independientes de tal manera que cada elemento de procesamiento pueda ejecutar su parte del algoritmo a la misma vez que los demás. Durante muchos años, la este tipo de computación se ha aplicado en máquinas de altas prestaciones, pero el interés en ella ha aumentado en la última década debido a las restricciones físicas que impiden el aumento en frecuencia. La computación paralela se ha convertido en el paradigma dominante en la arquitectura de computadores, principalmente con la aparición reciente de los procesadores multinúcleo. Los elementos de procesamiento pueden ser diversos e incluir recursos tales como un único ordenador con muchos procesadores y/o núcleos, varios ordenadores en red (*Grid computing*), hardware especializado o una combinación de los anteriores. Nos centraremos en el caso de la paralelización en procesadores multinúcleo, por ser la más sencilla de aplicar, y compatible con la mayoría de las máquinas actuales, incluyendo los ordenadores personales. La paralelización de un código pre-existente, requiere unos conocimientos de programación muy especializados, que escapan de los objetivos de esta Tesis. Sin embargo, existe la posibilidad de reemplazar etapas de cálculo susceptibles de paralelización, tales como funciones o subrutinas, por otras ya paralelizadas incluidas en librerías de dominio público o comerciales de alto rendimiento, alterando mínimamente el código fuente y, como veremos posteriormente, consiguiendo notables resultados.

Evidentemente, las etapas de cálculo que deben optimizarse tienen que ser aquellas que más tiempo consuman en el cálculo completo. Tras un análisis exhaustivo del código fuente, se llegó a la conclusión de que la operación matemática que más tiempo requería para el cálculo hidrodinámico riguroso de macromoléculas rígidas, representadas por nuestro modelos de esferas, es la inversión de la supermatriz de interacción hidrodinámica.

Existen diversos métodos matemáticos para llevar a cabo la inversión de una matriz. En la mayoría de ellos, es necesaria una factorización o descomposición previa a la inversión, que consume la mayor parte del tiempo. Una de las librerías más importantes especializadas en este tipo de cálculos es la denominada LAPACK (*Linear Algebra PACKage*), un software libre para cálculo numérico escrito en Fortran77, y que proporciona rutinas muy optimizadas para solucionar sistemas de ecuaciones lineales simultáneas, mínimos cuadrados, solución a sistemas de ecuaciones lineales, obtención de valores propios, problemas de valores singulares y factorizaciones e inversiones de matrices, entre otros. Esta librería está en continua evolución y ya hay disponibles numerosas versiones de la misma. Además, algunas casas comerciales ofrecen versiones propias y optimizadas de este conjunto de subrutinas.

Un aspecto muy importante cuando trabajamos con grandes matrices, además del tiempo de CPU, es la memoria que se necesita para almacenar dicha matriz, lo que limita el número máximo de elementos con los que podemos trabajar. Al ser una matriz simétrica, una vez construida, bastaría con almacenar uno de los dos triángulos que forman la misma, por encima o por debajo de la diagonal, con lo que ahorraríamos casi el 50% de la memoria disponible, ya que para una matriz de dimensión n ($n = 3N$ en nuestro caso), solo operaríamos con $n \times (n + 1)/2$ elementos. Muchas de las subrutinas o funciones disponibles para trabajar con matrices simétricas, incluyendo LAPACK, lo tienen en cuenta, y tenemos la opción de operar únicamente con uno de los dos triángulos junto a la diagonal. Esta forma de almacenar las matrices simétricas se denomina *almacenamiento empaquetado*. Los elementos de la matriz se almacenan consecutivamente en una variable de una dimensión, como si de un vector se tratara. El ahorro en memoria es considerable, aunque los cálculos por lo general sean más lentos. La otra opción sería trabajar con la matriz completa $n \times n$, en este caso, con n^2 elementos. Para tamaños de n muy grandes, la RAM consumida en almacenar la supermatriz es un factor a tener en cuenta, ya que puede presentarse el caso de que los recursos de memoria RAM en el equipo de cálculo sean insuficientes, obligando al sistema a utilizar medios más lentos (como el disco duro) para almacenar los datos, incrementándose el tiempo de cálculo en varios órdenes de magnitud, fracasando cualquier intento de optimización. Además, la memoria consumida en almacenar variables es memoria que deja de estar disponible para realizar las operaciones, obligando al sistema a reorganizar la memoria continuamente, perdiendo rendimiento. Algunos sistemas operativos no permiten direccionar variables en memoria que ocupen más de 2 GB, como ocurre en los sistemas operativos Windows y Linux de 32 bits, imposibilitando el cálculo a partir de un determinado número de elementos. Por estos motivos, además del tiempo de CPU requerido para la inversión, también es necesario tener en cuenta la memoria RAM que consume para realizar dicho cálculo.

Las librerías de funciones y subrutinas que vamos a evaluar, en un proceso de inversión de las típicas matrices que surgen en el cálculo de modelos de esferas, son las siguientes:

- Librería LAPACK 2.0 Publicada en 1993, y con el código fuente disponible en <http://www.netlib.org>, empleando la subrutina que implementa el método de factorización de Cholesky con almacenamiento empaquetado (SPPTRI). Este es el método que se ha ido utilizando previamente en los programas de cálculo hidrodinámico del

Grupo de Investigación, y los tiempos de CPU obtenidos serán tomados como referencia.

- Librería LAPACK 3.0 Publicada en 2006, y con el código fuente también disponible en <http://www.netlib.org>, presenta mejoras respecto a la anterior. Emplearemos dos métodos de inversión, (a) Factorización de Cholesky con almacenamiento empaquetado (SPPTRI), y (b) Factorización LU con almacenamiento normal (SGETRI).
- Librería LAPACK Intel® Math Kernel Library versión 10.0 (IMKL) Comercializada en 2008 (como complemento de la versión comercial del compilador Intel Fortran Compiler®), presenta optimizaciones específicas para el compilador y procesadores de la misma marca, que son los que principalmente utilizamos en nuestras compilaciones y cálculos. En este caso, el código fuente no está disponible, ya que viene incluida con el compilador, tanto para Linux como para Windows, en una versión comercial de bajo coste para instituciones académicas. De nuevo, emplearemos los dos métodos de inversión, (a) Factorización de Cholesky con almacenamiento empaquetado (SPPTRI) , y (b) Factorización LU con almacenamiento completo (SGETRI). También analizaremos la paralelización implementada de esta librería, lanzando el cálculo en multiproceso (2, 4 y 8 hilos).

Para llevar a cabo la prueba, mediante HYDROPRO se generaron modelos de esferas tipo concha (shell) para la proteína lisozima (código PDB 6LYZ). Se emplearon diez modelos compuestos por un número diferente de esferas N o elementos de fricción, desde $N = 155$ hasta $N = 1765$, lo cual supone un tamaño de matriz $n \times n$, con $n = 3N$, desde 465 hasta 5295, aumentando el grado de detalle del modelo a medida que aumenta el número de esferas. A continuación, se midieron los tiempos de CPU para cada modelo, empleando programas compilados y enlazados a las diversas librerías que vamos a comparar. Las pruebas de tiempo de CPU se realizaron en un equipo provisto de dos procesadores Xeon Quad E5440 2.83 GHz, con 8 GB de RAM compartida y sistema operativo Linux, el cual posee $2 \times 4 = 8$ núcleos (*cores*), por lo que los cálculos paralelizados pueden llevarse a cabo en hasta 8 núcleos o hilos de ejecución simultáneos, o en un número inferior opcionalmente. El programa fue compilado mediante el Intel Fortran Compiler 10.0 para Linux. Las pruebas se confirmaron posteriormente con resultados prácticamente idénticos, con el mismo compilador en un equipo con sistema operativo Windows XP.

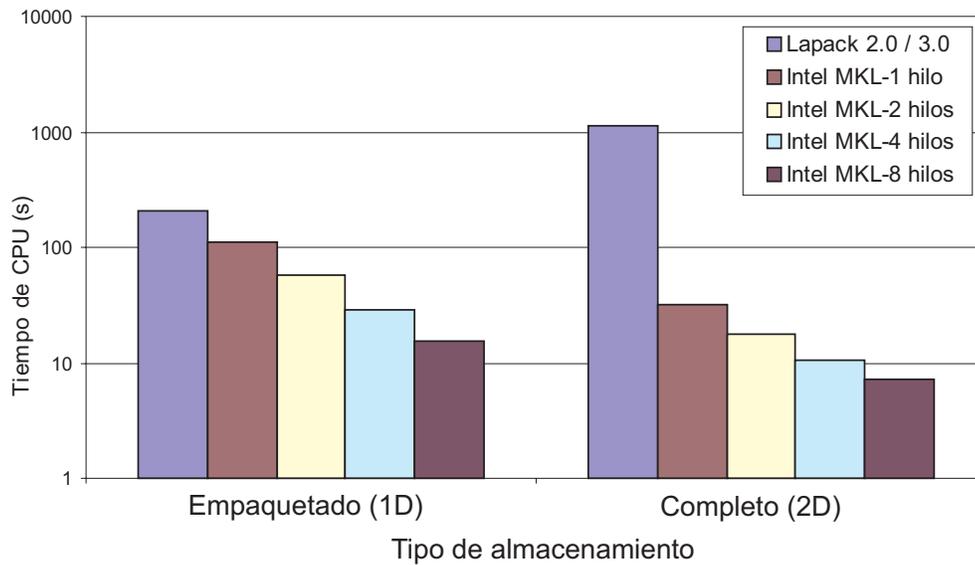


Figura III.1: Resultados de las pruebas de optimización del cálculo hidrodinámico. Tiempo de CPU para de las diversas subrutinas de inversión de matrices probadas. Se representa la suma de todos los tiempos de CPU, con $n = 465 - 5296$, con 1, 2, 4 y 8 hilos respectivamente.

Los resultados del estudio se muestran en la Tabla III.1 y Figuras III.1 y III.2, en las que se indica el tiempo de CPU (t_{CPU}) para los distintos procedimientos. El cálculo de HYDROPRO requiere la realización de toda una serie consecutiva de cálculos desde el menor al mayor N , por lo que el rendimiento viene dado por la suma de los tiempos de cálculo para cada N (recordando $n = 3N$). En primer lugar, podemos observar que el método con mayor rendimiento es el método de factorización de LU que nos ofrece la librería IMKL. El cálculo en monoproceso mediante esta librería (MKL 1x 2D) ya ofrece un rendimiento casi 7 veces superior (32.1 respecto a 208.5 segundos) al que ofrece el método que se ha ido empleando hasta ahora (LAPACK 2.0/3.0 1D). Si lanzamos el mismo cálculo en dos hilos (MKL 2x 2D), el rendimiento se multiplica en un factor de 2 respecto al cálculo monoproceso, y si lo lanzamos en 4 hilos, se aumenta casi 4 veces respecto al cálculo monoproceso (MKL 4x 2D), y así sucesivamente. No se observan diferencias entre el rendimiento que ofrecen las librerías de uso libre LAPACK 2.0 y 3.0. En la Tabla III.1 se muestran los resultados detallados, tanto para el tiempo totalizado como para cada n . Respecto al rendimiento de los métodos de almacenamiento empaquetado (1D) o completo (2D) en las diferentes versiones de LAPACK, en las librerías de Intel las que trabajan con

almacenamiento 2D son más rápidas que las equivalentes en almacenamiento 1D, en un factor de dos, aunque evidentemente consumen más memoria (el doble) en almacenar la matriz. No ocurre lo mismo en el caso de las librerías LAPACK 2.0/3.0 libres, donde las subrutinas de almacenamiento 1D son más veloces que las de almacenamiento 2D, aunque ya se ha demostrado que son las de menor rendimiento.

El método de factorización LU y almacenamiento completo en las librerías Intel MKL es el método más rápido visto hasta ahora en todas las pruebas realizadas. Además, incluye directivas (ya implementadas en ella) de paralelización, por lo que el rendimiento todavía es mucho mayor si aprovechamos esta característica. Tal vez, al ser un método general, aplicable a todas las matrices factorizables / invertibles, se han desarrollado más por parte del desarrollador que otros métodos. El único inconveniente que presenta es la alta ocupación de memoria, al almacenar la matriz completa, como comentamos anteriormente.

Independientemente de la metodología empleada, la inversión de matrices es un proceso computacionalmente costoso. El número de operaciones, y por tanto el tiempo de CPU (t_{CPU}), son proporcionales a n^3 , aunque esta dependencia solo se alcanza a n muy elevados, siendo el exponente algo inferior para los n moderados usados en nuestro trabajo. La dependencia (exponente) es casi la misma para los diversos algoritmos (LU, Cholesky, etc.), y por supuesto, para las diferentes implementaciones computacionales, como se aprecia en la Figura III.2. No obstante, nuestra actual implementación, LAPACK MKL 2D, respecto a la que se venía usando en el Grupo hasta ahora, LAPACK 2.0 1D, supone un incremento en velocidad (cociente de t_{CPU}) muy elevado, ya en un simple equipo de 2 “cores”, como son los ordenadores personales más sencillos, hasta una estación de trabajo de 8 “cores” repartidos en dos procesadores “Quad core”. Estos factores de eficacia se reflejan en la Figura III.2b.

Influencia de la arquitectura y precisión en el cálculo hidrodinámico

Aunque el modo de cálculo habitual en las ejecuciones de nuestros programas de la familia HYDRO ha sido de precisión simple (SP), que consiste en emplear enteros y reales de 32 bits, en la publicación del programa HYDROPRO en el año 2000 [12] se menciona la necesidad de utilizar la doble precisión (variables de 64 bits) en el cálculo hidrodinámico a partir de un determinado número de elementos de fricción N , frente a la utilización de variables de precisión simple. Según este trabajo, mientras que los coeficientes rotacionales y traslacionales no varían significativamente por debajo de $N \approx 3000$ elementos,

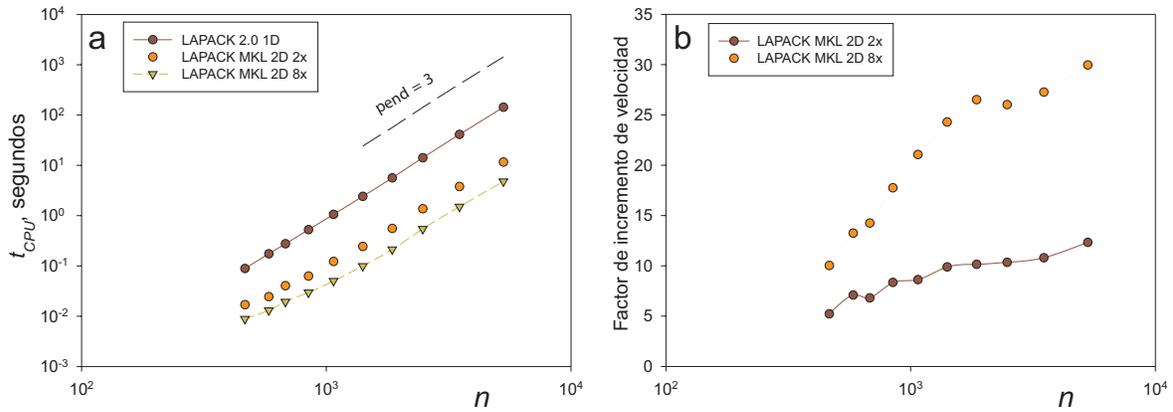


Figura III.2: (a) Variación con n del t_{CPU} de diversos procedimientos. (b) Incremento de rendimiento de los procesos actualmente implementados frente al anterior (véase texto).

N	Librería									
	LAPACK 2.0 / 3.0		LAPACK MKL 1x		LAPACK MKL 2x		LAPACK MKL 4x		LAPACK MKL 8x	
	1D	2D	1D	2D	1D	2D	1D	2D	1D	2D
465	0.089	0.538	0.028	0.022	0.025	0.017	0.026	0.011	0.029	0.009
582	0.174	1.042	0.047	0.034	0.046	0.024	0.045	0.017	0.046	0.013
681	0.276	1.655	0.075	0.059	0.073	0.040	0.071	0.026	0.070	0.019
846	0.526	3.140	0.136	0.094	0.122	0.063	0.128	0.040	0.127	0.030
1071	1.061	6.316	0.269	0.202	0.191	0.123	0.181	0.075	0.203	0.050
1410	2.414	14.401	0.590	0.439	0.395	0.244	0.194	0.142	0.205	0.099
1860	5.640	33.246	1.632	0.988	0.973	0.556	0.507	0.308	0.285	0.213
2478	14.150	78.859	6.236	2.388	3.255	1.367	1.658	0.779	0.918	0.544
3501	41.062	221.412	21.797	6.539	11.166	3.800	5.787	2.199	3.044	1.506
5295	143.206	765.137	80.399	21.317	40.482	11.611	20.488	6.892	10.658	4.780
Suma	208.597	1125.745	111.208	32.081	56.730	17.845	29.087	10.490	15.584	7.262

Tabla III.1: Tiempo de CPU (s) de las diversas subrutinas de inversión de matrices para varios N . (Abreviaturas: *MKL* es la librería LAPACK de Intel. 1D es el método de descomposición de Cholesky (SPPTRF) con almacenamiento empaquetado, y 2D es el método de factorización de LU (SGETRI) con almacenamiento completo. 1x, 2x, 4x y 8x son los cálculos multi-hilo con 1, 2, 4, 8 hilos simultáneos respectivamente.)

la viscosidad intrínseca no sufre variación significativa si se trabaja por debajo de los $N \approx 1000$ elementos. El empleo de variables de precisión simple como norma general, dejando el cálculo a precisión doble para casos puntuales, estaba justificado debido al alto coste computacional que suponía trabajar con variables de precisión doble en máquinas ya obsoletas actualmente con arquitecturas de 32 bits, que requerían el doble de tiempo frente al mismo cálculo de precisión sencilla.

Respecto a la elección de la arquitectura o plataforma, tradicionalmente se han ido empleando en el Grupo programas compilados a 32 bits. Actualmente, las máquinas con procesadores de 64 bits se han extendido hasta convertirse en equipos habituales de cálculo. En este tipo de procesadores, las variables de datos de doble precisión, tanto enteras como reales, son procesadas de una sola vez, ya que el tamaño de palabra del procesador coincide con el de la variable, por lo que ya no supone una pérdida de rendimiento emplear este tipo de datos. Además, debido a que los compiladores y librerías ya incluyen optimizaciones exclusivas para este tipo de procesadores, es de esperar que los cálculos realizados en la arquitectura de 64 bits sean más rápidos que sus equivalentes en 32 bits, tanto a precisión simple como doble.

En esta sección estudiaremos el efecto de la precisión y la arquitectura en el cálculo hidrodinámico, analizando los valores de los coeficientes traslacionales y la viscosidad intrínseca, y su rendimiento en el tiempo, para diversos n , en ambas arquitecturas y en ambos tamaños de variable. Los modelos de esferas utilizados fueron los mismos que en el apartado anterior, diez modelos de concha de la proteína lisozima compuestos por un número diferente de esferas. Para la inversión de la supermatriz de interacción hidrodinámica se utilizaron las subrutinas SPTRF + SPTRI para precisión simple y DPPTRF + DPPTRI para precisión doble, incluidas en la librería LAPACK 3.2 de acceso libre, obtenidas de www.netlib.org. Los cálculos, y los tiempos se obtuvieron en una ejecución monohilo, en un computador Intel Xeon Quad E5440 2.83 GHz, con 8 GB de RAM compartida y sistema operativo Linux. Los resultados se muestran en la Tabla III.2 y Figura III.3.

Como conclusión, no se observan diferencias mayores del 5% en las propiedades en disolución calculadas a precisión simple respecto a doble. Observamos que que la DP no modifica los resultados obtenidos a SP en el caso del coeficiente de difusión traslacional D_t (diferencias menores del 1%). Sin embargo en el caso de la viscosidad intrínseca $[\eta]$, para N superiores a 1500 ($n > 5000$), ya se notan diferencias del orden del 4%. Es en este rango donde comienza la necesidad donde podría considerarse la conveniencia de utilizar

t_{CPU}				
n	32SP	64SP	32DP	64DP
465	0.055	0.062 / 1.127*	0.156 / 2.836*	0.054 / 0.982*
582	0.103	0.118 / 1.146*	0.298 / 2.893*	0.100 / 0.971*
681	0.161	0.185 / 1.149*	0.470 / 2.919*	0.156 / 0.969*
846	0.300	0.348 / 1.160*	0.887 / 2.957*	0.289 / 0.963*
1071	0.592	0.695 / 1.174*	1.777 / 3.002*	0.574 / 0.970*
1410	1.319	1.555 / 1.179*	4.240 / 3.215*	1.950 / 1.478 *
1860	3.316	3.781 / 1.140*	10.577 / 3.190*	7.320 / 2.207*
2478	10.495	11.199 / 1.067*	25.813 / 2.460*	20.700 / 1.972*
3501	34.417	35.555 / 1.033*	74.174 / 2.155*	62.800 / 1.825*
5295	125.070	128.076 / 1.024*	258.057 / 2.063*	224.000 / 1.791*
Suma	175.83	181.570 / 1.033*	376.450 / 2.141*	317.940 / 1.808*

$D_t \times 10^6 \text{ cm}^2/\text{s}$				
n	32SP	64SP	32DP	64DP
465	1.108	1.108 / 1.000*	1.108 / 1.000*	1.110 / 1.002*
582	1.094	1.094 / 1.000*	1.094 / 1.000*	1.090 / 0.997*
681	1.100	1.100 / 1.000*	1.100 / 1.000*	1.100 / 1.000 *
846	1.102	1.102 / 1.000*	1.102 / 1.000*	1.100 / 0.998*
1071	1.100	1.100 / 1.000*	1.100 / 1.000*	1.100 / 1.000*
1410	1.095	1.095 / 1.000 *	1.095 / 1.000*	1.090 / 0.995 *
1860	1.090	1.090 / 1.000*	1.090 / 1.000*	1.090 / 1.000*
2478	1.091	1.091 / 1.000*	1.090 / 0.999*	1.090 / 0.999*
3501	1.089	1.089 / 1.000*	1.087 / 0.999*	1.090 / 1.001*
5295	1.092	1.092 / 1.000*	1.089 / 0.997*	1.090 / 0.998*

$[\eta] \text{ cm}^3/\text{g}$				
n	32SP	64SP	32DP	64DP
465	3.600	3.600 / 1.000*	3.600 / 1.000*	3.600 / 1.000*
582	3.729	3.730 / 1.000*	3.729 / 1.000*	3.730 / 1.000*
681	3.640	3.641 / 1.000*	3.640 / 1.000*	3.640 / 1.000 *
846	3.601	3.602 / 1.000*	3.602 / 1.000*	3.600 / 1.000*
1071	3.627	3.628 / 1.000*	3.628 / 1.000 *	3.630 / 1.001*
1410	3.640	3.640 / 1.000*	3.639 / 1.000 *	3.640 / 1.000*
1860	3.674	3.674 / 1.000*	3.670 / 0.999*	3.670 / 0.999*
2478	3.634	3.634 / 1.000*	3.625 / 0.998*	3.630 / 0.999*
3501	3.652	3.652 / 1.000*	3.622 / 0.992*	3.620 / 0.991*
5295	3.737	3.736 / 1.000*	3.585 / 0.960*	3.590 / 0.961*

Tabla III.2: Tiempo de CPU, coeficiente de difusión y viscosidad intrínseca del modelo de concha de lisozima para diversos n , plataforma y tamaño de variable. *Valores referidos al resultado obtenido en 32SP, tomado como referencia.

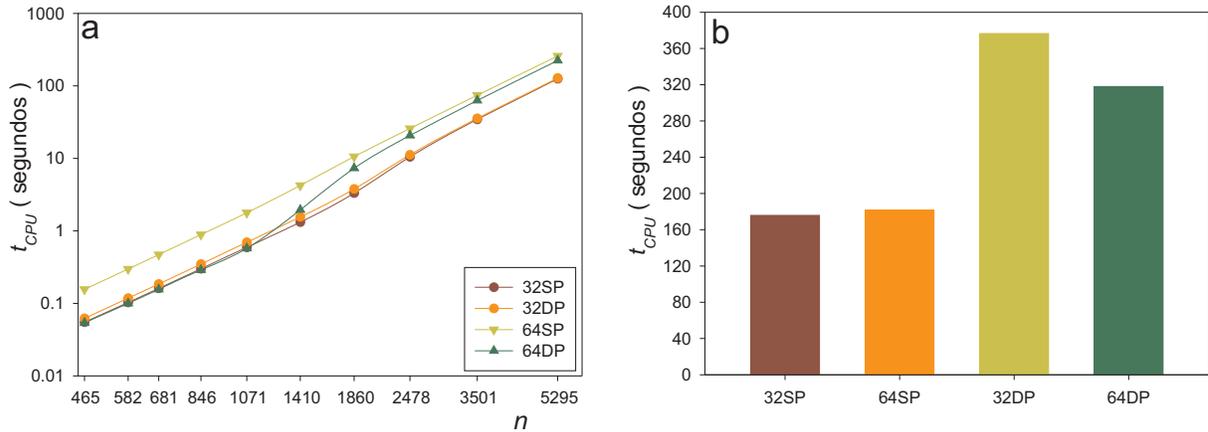


Figura III.3: (a) Variación con n del t_{CPU} de las diversas arquitecturas y tamaño de variable. (b) Suma de los t_{CPU} para todos los n probados.

doble precisión para el cálculo de $[\eta]$. Nuestro criterio es que no está justificado emplear modelos hidrodinámicos con un número excesivo (mucho mayor que 2000) de elementos. El aumento de precisión o detalle que se consigue en el modelo es infructuoso, ya que, (a) las teorías hidrodinámicas aún tienen algunas imperfecciones, (b) las teorías tienen una resolución modesta (son insensibles a detalles muy finos), y (c) los valores experimentales tienen errores en el rango 1-5%, mayor que el del refinamiento computacional de N tan altos. Por tanto, adoptamos el modelo de cálculo SP.

Respecto a la elección de la arquitectura, los cálculos realizados a 64 bits a precisión simple se retrasan ligeramente (un 3%) respecto al mismo cálculo a 32 bits. Sin embargo, a doble precisión, sí se observa un incremento notable en rendimiento a 64 respecto a 32 bits, y por debajo de $n = 1000$, supera en rendimiento a los otros tres casos.

III.1.2. Interfaces gráficas para los programas HYDRO++ e HYDROPRO

Hasta el momento, todos los programas desarrollados en el grupo de investigación están basados en unas interfaces de texto, donde los datos se introducen en un archivo de texto ASCII y los resultados se imprimen en otro archivo de texto. Aunque los programas son perfectamente funcionales, requieren unos mínimos conocimientos de informática, que en ocasiones no están presentes en el usuario final, sobre todo en el ámbito científico o docente, más familiarizado con un entorno de ventanas que con consolas de MS-DOS o GNU/Linux. En la mayoría de los casos, utilizar alguno de estos programas por primera vez requiere una lectura profunda del manual de usuario correspondiente, y la mayoría de errores que se producen son debidos a trivialidades como la elección errónea del separador decimal o el uso de unidades físicas diferentes a las que trabajan los programas. Mediante una interfaz gráfica se evita que el usuario cometa estos errores y permite la visualización directa, sin programas externos, de las estructuras macromoleculares con las que se trabaja. Además, facilitando el uso de nuestros programas, se consigue más difusión y reconocimiento por parte de los usuarios, y de la comunidad científica en general.

La interfaz gráfica de usuario (*Graphical User Interface, GUI*) es un programa informático que actúa como intermediario entre un programa y el usuario, utilizando un conjunto de imágenes y objetos gráficos para representar la información y acciones disponibles en la interfaz. Su principal uso consiste en proporcionar un entorno visual sencillo para permitir la comunicación con el programa, que posibilita, a través del uso y la representación del lenguaje visual, una interacción amigable con un programa o sistema informático. Desde el punto de vista del programador, una de las principales ventajas que presentan es que nos permite crear programas que realicen de la misma forma las tareas más frecuentes, como cargar o guardar un archivo, porque la interfaz proporciona mecanismos estándar de control como ventanas y cuadros de diálogo. Esto hace que el programador pueda concentrarse en la funcionalidad, ya que no está sujeto a los detalles de la visualización ni a la entrada a través del ratón o del teclado.

Hemos construido interfaces gráficas para dos de nuestros programas más utilizados por la comunidad científica, como son los programas HYDRO++ e HYDROPRO, empleando el lenguaje de programación Visual Basic .NET 2005, incluido en la distribución del paquete Microsoft Visual Studio 2005[®]. Estas interfaces se apoyan en los ejecutables de consola de HYDRO++ e HYDROPRO, y se encargan principalmente de generar los archivos de entrada de ambos, y posteriormente, mostrar por pantalla los resultados. Nótese que la ejecución

propriadamente dicha es realizada por los ejecutables de consola. Todo el proceso puede ahora realizarse de manera visual e interactiva e incluye avisos cuando algún dato está mal introducido.

En la Figura III.4 podemos observar las capturas de pantalla de ambas interfaces gráficas. En la parte superior, que corresponde a la interfaz del programa HYDROPRO (que hemos denominado WinHYDROPRO), observamos a la izquierda la ventana principal, donde se carga la estructura atómica (archivo PDB) y se introducen los datos correspondientes a las magnitudes físicas del disolvente (T , η_0 , ρ), del soluto (M , \bar{v}), y los radios σ y el número de ellos para el cálculo hidrodinámico del modelo de concha, además de otras opciones, como el calculo de $P(r)$, $P(h)$, etc... Una vez realizado el cálculo, podemos visualizar el modelo atómico y primario y el archivo de resultados (derecha). En la parte inferior de la Figura III.4, podemos observar la captura de pantalla de la interfaz del programa HYDRO++ (que hemos denominado WinHYDRO++). De nuevo, a la izquierda, la ventana principal, donde se introducen las magnitudes físicas del disolvente y soluto, y a la derecha la visualización del modelo de esferas, que puede ser cargado de un archivo ya existente, o construirlo, ya que se incluye la posibilidad de construir modelos de esferas, con la posibilidad de guardarlas y recuperarlas posteriormente.

Pese a que la realización de estas interfaces gráficas para Windows es un avance importante, todavía quedan tareas por realizar, como la posibilidad de construir estas mismas interfaces para otras plataformas, como GNU/Linux o Mac, o incluir en la misma interfaz el propio cálculo hidrodinámico. También sería deseable la construcción de interfaces para otros programas como HYDRONMR, Single-Multi-HYDFIT o el paquete Simuflex.

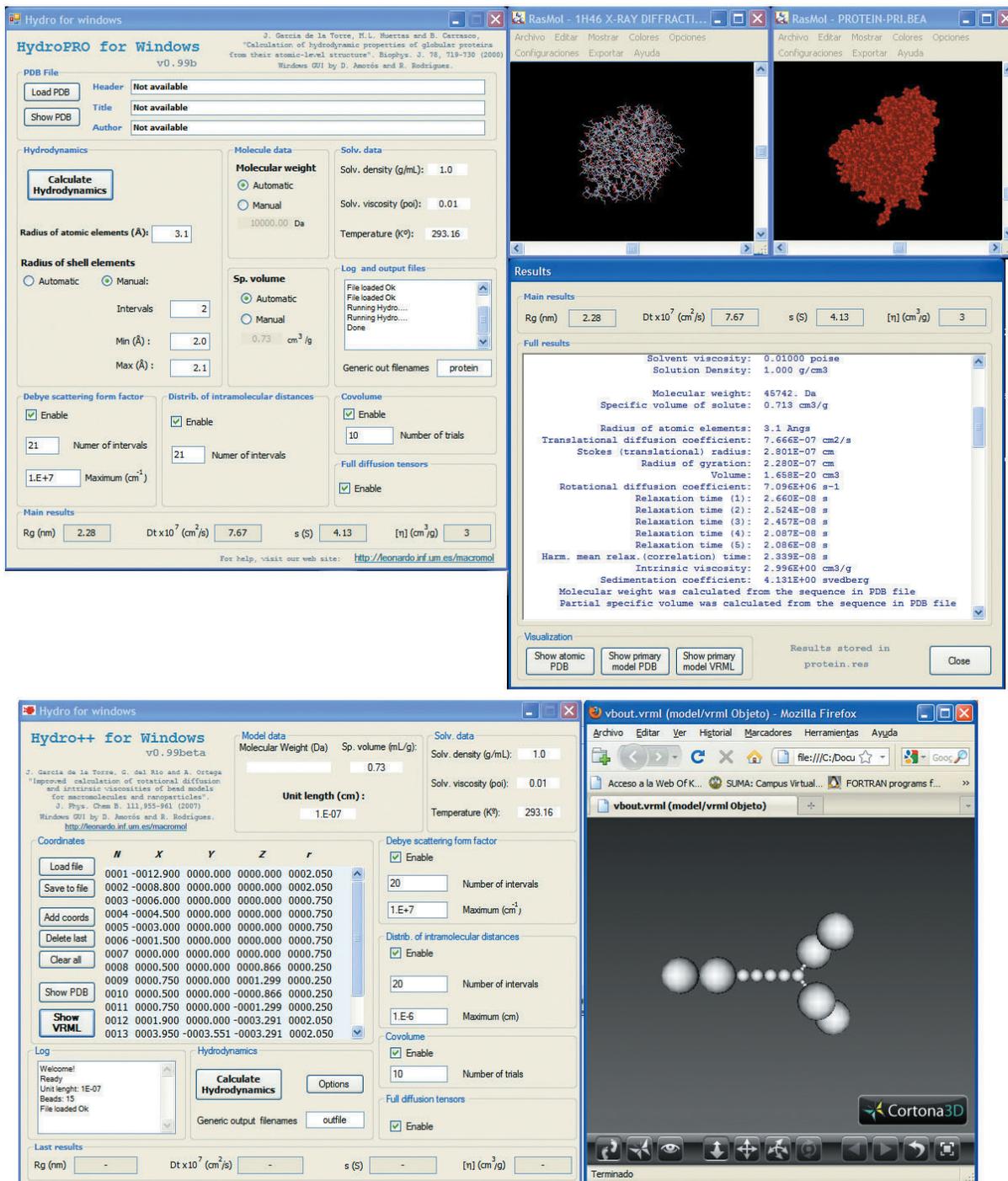


Figura III.4: Capturas de pantalla de los programas WinHYDROPRO (superior) y WinHYDRO++ (inferior). Más detalles en el texto.

III.2. Mejoras en el tratamiento de la viscosidad

Mientras que en el campo experimental, las técnicas de determinación de la viscosidad intrínseca, $[\eta]$, han avanzado a pasos agigantados en los últimos tiempos [36, 37, 38, 39, 40], la descripción teórica y la predicción computacional en modelos que representan partículas con forma arbitraria continúa presentando numerosas dificultades. La teoría clásica de modelos de esferas de Kirkwood y Riseman (KR) [41, 42, 43] fue diseñada para modelos que representan varillas rígidas o flexibles y compuestas por un número N de esferas idénticas. En estos modelos, el diámetro de las esferas es mucho menor que el tamaño total de la partícula, por lo que se podría aceptar que los puntos de fricción del modelo se encuentran en el centro de dichas esferas. Posteriormente, García de la Torre y Bloomfield [44] consideraron el tamaño finito de las esferas representando las interacciones hidrodinámicas mediante el tensor de Oseen, y fue implementado en la primera versión de HYDRO [22]. Su resultado para la viscosidad intrínseca [8, 45], que denominaremos en lo sucesivo *sin corrección* o $[\eta]_{KRM}$, presenta problemas en modelos compuestos por pocas esferas, por lo que se presentó diez años después una primera corrección relacionada con el volumen de la partícula [46] basada en la extensión de la teoría KR considerando el efecto del tamaño finito de las esferas, llegando a la conclusión de que el valor correcto de la teoría KR debería haber sido

$$[\eta]_{KRV} = [\eta]_{KRM} + \frac{10\pi N_A}{3M} \sum_{i=1}^N \sigma_i^3 = [\eta]_{KRM} + \frac{5N_A V}{2M} \quad (\text{III.1})$$

donde M es la masa molecular, σ_i es el radio de la esfera i , y $V = (4\pi/3) \sum \sigma_i^3$ es el volumen total del conjunto de esferas. El subíndice KRV indica que se trata de una corrección determinado por el volumen de la partícula, que denominamos corrección *completa*. El defecto que presenta la teoría KR de considerar cada elemento como un solo punto de fricción, puede ser subsanado, alternativamente, mediante una estrategia de modelado, que consiste en la sustitución de las esferas por un cubo formado por varias *mini-esferas* [47, 48]. En la denominada *sustitución cúbica*, la fuerza de fricción en este caso no se concentra en un único punto, sino que se distribuye alrededor de ocho centros, cercanos a la superficie de la esfera que se reemplaza. El número total de esferas se incrementa en un factor de 8, con el inconveniente de que el tiempo de cálculo se incrementa en un factor de $8^3 = 512$. En los casos donde se conocen los valores exactos de viscosidad, mediante teorías de la mecánica de fluidos, los resultados obtenidos de la sustitución cúbica son

prácticamente exactos. Los valores de viscosidad exactos obtenidos por este método, se tomaran como referencia y se indicarán como $[\eta]_{REF}$ en lo sucesivo. En un trabajo del año 1998 [46], se presentaron los resultados de la corrección de volumen para algunas estructuras, llegando a la conclusión de que los valores de $[\eta]_{KRV}$ son generalmente más cercanos a los exactos que los valores sin corregir $[\eta]_{KRM}$, aunque el valor exacto suele estar situado entre ambos. En un trabajo reciente [23] se sugiere la posibilidad de desarrollar una corrección de volumen empírica, fraccional o intermedia, a partir de estos dos valores, definiéndose un nuevo valor de viscosidad que denominaremos $[\eta]_{KRI}$:

$$[\eta]_{KRI} = [\eta]_{KRM} + f_\eta \frac{5N_A V}{2M} \quad (\text{III.2})$$

donde $0 \leq f_\eta \leq 1$. Para que $[\eta]_{KRI}$ coincidiera con el valor exacto, $[\eta]_{REF}$, sería:

$$f_\eta = \frac{[\eta]_{REF} - [\eta]_{KRM}}{[\eta]_{KRV} - [\eta]_{KRM}} = \frac{[\eta]_{REF} - [\eta]_{KRM}}{(5N_A V/2M)} \quad (\text{III.3})$$

Nuestro propósito consiste en desarrollar un método sencillo para estimar f_η a partir de características geométricas del modelo hidrodinámico. Como se comentó en un trabajo previo [46], en el caso de estructuras donde una esfera de gran tamaño esté rodeada de otras más pequeñas, la corrección necesaria está muy cercana a la completa, con un factor $f_\eta \approx 1$. Por consiguiente, es necesario evaluar la ‘fragmentación’ del modelo. El método que se sugiere es el siguiente: Supongamos que las esferas de nuestro modelo se ordenan de mayor a menor radio, siendo $i=1$ la mayor, $i=2$ la segunda, e $i = N$ por consiguiente la más pequeña. Llamaremos ϕ_i a la fracción del volumen total que ocupa la esfera i . Definimos la suma acumulativa $S_i = \sum_{k=1}^i \phi_k$, donde $S_0 = 0$, $S_1 = \phi_1$ y $S_N = 1$. Entonces, definimos una nueva función

$$S = N - \frac{1}{2} - \sum_{i=1}^N \frac{S_i + S_{i-1}}{2} \equiv N - \sum_{i=1}^N S_i \quad (\text{III.4})$$

que presenta las siguientes propiedades: (a) En un modelo compuesto por esferas idénticas, S toma el mayor valor posible $S = N/2 - 1/2 = (N - 1)/2$. (b) En un modelo donde una esfera es la dominante respecto a las demás, $\phi_1 \approx 1$, $S \approx 0$. Además, $S = 0$ para $N = 1$, por ejemplo, en el caso de modelos compuestos por una sola esfera.

Otro aspecto a tener en cuenta es que en los casos donde las esferas son de idéntico tamaño, el factor f_η necesario para obtener resultados correctos depende de la forma global del modelo, particularmente de la mayor o menor elongación de la estructura. En estructuras globulares muy isométricas, f_η es cercano a 0, mientras que se precisan valores altos de f_η en el caso de cadenas elongadas de esferas. Es necesario un indicador de la anisometría de la partícula. Una posibilidad, que fue la elegida en [46], sería basar el indicador en el tensor de difusión D_{tt} , que es uno de los resultados suministrados por el cálculo completo de propiedades hidrodinámicas en las metodologías de los programas de la serie HYDRO. Sin embargo, hay otros procedimientos para evaluar la viscosidad intrínseca [49, 50] que no requieren dicho tensor. Por ello, ahora representamos la anesfericidad del modelo con un criterio geométrico, no hidrodinámico. Este procedimiento está basado en el tensor de giro \mathbf{G} , que se calcula, para un conjunto de N esferas no-solapantes:

$$\mathbf{G} = \sum_{i=1}^N f_i \left(s_i^\alpha s_i^\beta + (3/5)\sigma_i^2 \delta_{\alpha\beta} \right) \quad (\text{III.5})$$

donde \mathbf{s}_i es el vector de posición del elemento i respecto al centro de masas, $\mathbf{s}_i = \mathbf{r}_i - \mathbf{r}_{CM}$, y \mathbf{r} es el vector de posición respecto a un origen arbitrario. α, β son los componentes del vector ($\alpha, \beta = x, y, z$). El segundo término, es el radio de giro de las esferas individuales, equivalente a $(3/5)\sigma_i^2 \delta_{\alpha\beta}$, donde σ_i es el radio de la esfera y $\delta_{\alpha,\beta}$ es el delta de Kronecker's, se omite en el caso de modelos compuestos por esferas idénticas y relativamente pequeñas. El centro de masas se expresa como:

$$\mathbf{r}_{CM} = \sum_{i=1}^N f_i \mathbf{r}_i \quad (\text{III.6})$$

donde

$$f_i = \sigma_i^3 / \sum_i \sigma_i^3 \quad (\text{III.7})$$

es la fracción de volumen del elemento i . Si G_1, G_2, G_3 son los tres valores propios de \mathbf{G} , la anesfericidad se define como [51]:

$$A_G = \frac{(G_2 - G_1)^2 + (G_3 - G_1)^2 + (G_3 - G_2)^2}{2(G_1 + G_2 + G_3)^2} \quad (\text{III.8})$$

donde en el caso de partículas isométricas, no necesariamente esféricas, $G_1 = G_2 = G_3$, y $A_G = 0$. Si la partícula es muy elongada, $G_1 \gg G_2, G_3$ el resultado es $A_G = 1$. También podemos, alternativamente, emplear la cantidad A_D , definida análogamente en términos de valores propios D_1, D_2 y D_3 del tensor de difusión traslacional.

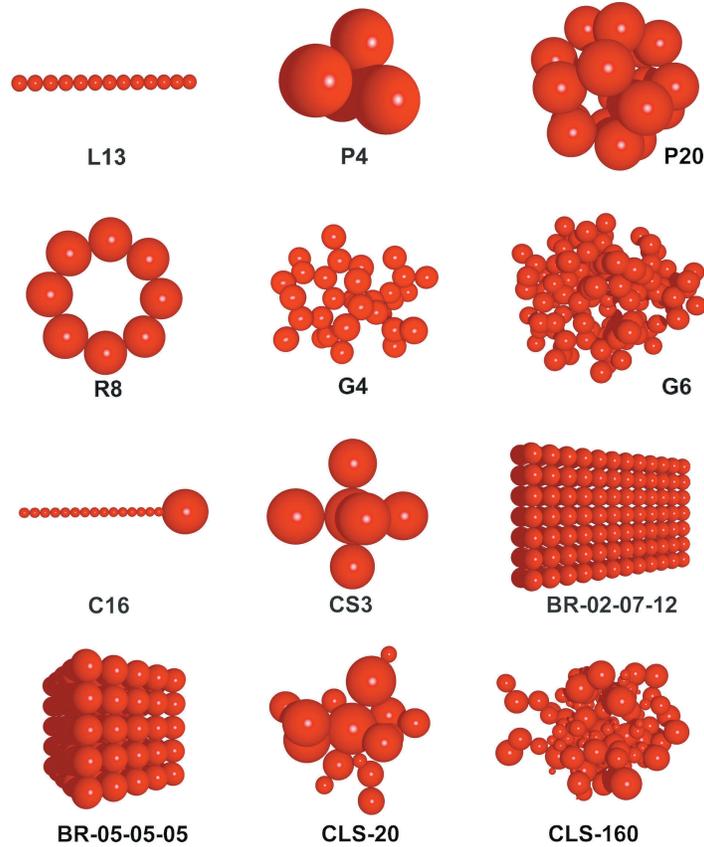


Figura III.5: Selección de estructuras utilizadas en el estudio. La codificación empleada se describe en [23] o en el texto.

Debemos encontrar una forma de la función $f_\eta(S, A; p_1, p_2, p_3, \dots)$ que refleje los efectos de la fragmentación S y de la forma de la partícula, mediante A_D o A_G en los modelos de esferas. Esta función será utilizada para determinar los valores exactos de viscosidad a partir de $[\eta]_{KRM}$ y V , obteniendo una corrección intermedia $[\eta]_{KRI}$ empleando la ecuación III.2.

Como ocurría en la corrección de volumen anterior, esta función tendrá parámetros o coeficientes $P_1, P_2 \dots$ ajustables, procediendo de una manera semi-empírica. Para un con-

junto variado de estructuras (ver a continuación) hemos evaluado la viscosidad intrínseca por tres medios, obteniendo $[\eta]_{KMR}$, $[\eta]_{KRV}$ y $[\eta]_{REF}$. Para cada estructura, hemos obtenido f_η a partir de $[\eta]_{KMR}$ y $[\eta]_{REF}$ empleando la ecuación III.3. En la determinación de la función y sus parámetros óptimos se trató de minimizar las desviaciones de los valores de $[\eta]_{KRI}$ calculados con la nueva corrección intermedia (ecuación III.2) respecto a los valores exactos $[\eta]_{REF}$. Para cada estructura se evaluó la desviación porcentual

$$\Delta \% = 100 \left[\frac{[\eta]_{KRI} - [\eta]_{REF}}{[\eta]_{REF}} \right] \quad (\text{III.9})$$

y para el conjunto de estructuras (N_{str}) se adoptó un criterio mínimo cuadrático, basado en una desviación global

$$\Delta \%_{global} = \sqrt{\frac{1}{N_{str}} \sum_i^{N_{str}} (\Delta \%_i)^2} \quad (\text{III.10})$$

La estimación empírica de f_η mediante una función que dependa de S y A_G/A_D se ha realizado a partir de varios tipos de estructuras, de un total de 47. La mayoría de ellas aparecen descritas en el trabajo que precede a este [23]: Dímeros, estructuras lineales, poligonales y poliédricas de esferas iguales y desiguales. Algunos modelos representan dendrímeros, paralelepípedos y “clusters”, otros tienen formas peculiares. El conjunto de modelos presenta una gran variedad en cuanto a los dos factores que hemos considerado primordiales: la fragmentación y la anisometría. También incluimos modelos ‘macizos’ de proteínas globulares que emplean otros autores. Los modelos de concha de biomoléculas con un gran número de esferas pequeñas e idénticas, prácticamente no precisan ningún tipo de corrección en este sentido, por lo que no se han incluido. Una selección de estas estructuras puede verse en la Figura III.5.

Resultados

En la Tabla III.3 se muestran los resultados obtenidos. Los valores de referencia $[\eta]_{REF}$ están normalizados al valor que presenta una esfera de igual volumen $[\eta]_{REF}^* = [\eta]_{REF}/(5N_A V/2M)$. Nótese que de esta manera no depende del tamaño absoluto. Para los valores de KRM, KRV y KRI se muestra los correspondientes $\Delta \%$ de cada estructura y el el global $\Delta \%_{global}$.

Para hallar la función $f_\eta(S, A; p_1, p_2, p_3, \dots)$ se probaron diversos tipos de ecuaciones que cumplieran con ciertos requisitos en su dependencia con las variables S y A . Respecto al factor de anesfericidad A se empleó tanto A_D como A_G , obteniéndose ajustes ligeramente mejores con el factor A_G , además de que su cálculo es más rápido. Se emplearon métodos iterativos, en concreto la herramienta SOLVER de Microsoft Excel® para evaluar y ajustar los parámetros de las diversas ecuaciones que se probaron. Como resultado del ajuste, obtuvimos una ecuación simple pero efectiva:

$$f_\eta = 1 - 0.8474 \frac{S}{S+1} + 1.0171 A_G \frac{S}{S+1} - 0.3801 A_G \quad (\text{III.11})$$

empleando $S/(S+1)$ en lugar de S para que su rango varíe de 0 a 1, como ocurre con A_G . En la Figura III.6 puede verse una representación de la función f_η . Esta ecuación permite reducir el error promedio de las 47 estructuras hasta el 4.3%, frente al 35.8% en ausencia de cualquier tipo de corrección (KRM), y al 24.0% con la corrección completa (KRV), y al 8% obtenido en una versión anterior de la corrección intermedia anterior [23]. Este error está muy próximo o por debajo del obtenido experimentalmente en nuestros días, por lo que los valores de $[\eta]$ obtenidos empleando esta corrección, pueden ser considerados prácticamente exactos.

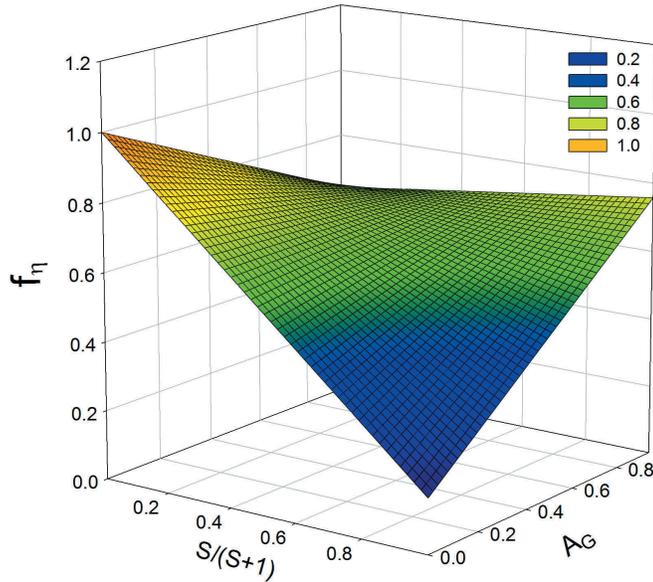


Figura III.6: Representación gráfica tridimensional de la ecuación f_η (ec. III.11).

III.2. MEJORAS EN EL TRATAMIENTO DE LA VISCOSIDAD

Tabla III.3: Resultados de la estimación empírica de f_η para los modelos empleados en el estudio. Los códigos están descritos en [23] o en el texto.

Estructura	N	REF ^a	$\Delta\%(KRM)$	$\Delta\%(KRV)$	A_G	S	f_η (obs)	f_η (cal)	$\Delta\%(KRI)$
M1	1	1.78	-100	0	0.000	0.000	1.000	1.000	0.0
D1	2	2.458	-53.1	20.1	0.391	0.500	0.726	0.701	-1.8
D2	2	6.348	-68.1	17	0.083	0.111	0.801	0.892	7.8
D3	2	13.5	-80.4	13	0.009	0.036	0.861	0.968	9.9
L3	3	3.292	-40.1	14.6	0.6664	1.0000	0.7328	0.662	-3.88
L4	4	4.294	-30.9	11.1	0.7972	1.5000	0.7361	0.675	-2.56
L6	6	6.608	-20.2	7.1	0.9046	2.5000	0.7396	0.708	-0.86
L8	8	9.44	-15	4.1	0.9452	3.5000	0.7856	0.729	-1.07
L10	10	12.69	-11.7	2.4	0.9646	4.5000	0.8278	0.743	-1.21
L13	13	18.31	-8.8	1	0.9789	6.0000	0.8944	0.755	-1.37
L20	20	34.75	-5.5	-0.3	0.9910	9.5000	1.0667	0.769	-1.54
P4	4	2.542	-30	40.8	0.0008	1.5000	0.4233	0.492	4.84
P6	6	2.443	-24.3	49.4	0.0000	2.5000	0.3294	0.395	4.81
P8	8	2.803	-24.1	40.1	0.0000	3.5000	0.3750	0.341	-2.19
P12	12	3.876	-22.7	47	0.0000	5.5000	0.3259	0.283	-2.99
P20	20	3.632	-17.2	32.4	0.0000	9.5000	0.3472	0.233	-5.65
R3	3	2.514	-37.5	34	0.1189	1.0000	0.5244	0.592	4.81
R4	4	2.791	-33.2	31.3	0.1479	1.5000	0.5150	0.526	0.68
R6	6	3.457	-27	25.1	0.1890	2.5000	0.5183	0.460	-3.03
R8	8	4.422	-24.5	16.2	0.2112	3.5000	0.6028	0.428	-7.13
G3	15	6.812	-13.1	13.3	0.1903	7.0000	0.4972	0.356	-3.74
G4	31	7.936	-9.9	12.8	0.1211	15.0000	0.4367	0.275	-3.67
G5	63	7.97	-8.2	14.4	0.0791	31.0000	0.3639	0.227	-3.09
G6	127	7.101	-7	18.4	0.0454	63.0000	0.2744	0.194	-2.04
C2	2	1.828	-90.2	8.3	0.0003	0.0079	0.9156	0.993	7.65
C4	4	2.139	-75.1	9.1	0.0089	0.0469	0.8922	0.959	5.63
C6	6	2.66	-59.6	8.1	0.0471	0.1154	0.8806	0.899	1.27
C11	11	4.884	-32.7	4.1	0.3158	0.4074	0.8878	0.728	-5.90
C16	16	8.424	-19.5	1.9	0.6001	0.8571	0.9133	0.663	-5.36
A1	7	3.531	-21.1	29.9	0.2067	1.5574	0.4139	0.533	6.09
A2	7	3.046	-32.6	26.5	0.4252	0.6517	0.5522	0.675	7.24
A3	7	3.558	-20	30.6	0.0498	2.3168	0.3956	0.425	1.47
B1	7	6.557	-20.3	7.1	0.8680	1.5574	0.7411	0.692	-1.36
B2	7	3.787	-34.5	13	0.6206	0.6517	0.7256	0.679	-2.22
B3	7	7.645	-17.4	6.1	0.9185	2.3168	0.7394	0.712	-0.66
BR-02-04-08	64	2.672	-11	26.5	0.4411	31.5000	0.2934	0.446	5.71
BR-02-07-12	168	3.451	-8.1	20.9	0.3997	83.5000	0.2787	0.412	3.88
BR-03-03-03	27	1.76	-16.2	40.6	0.0000	13.0000	0.2858	0.213	-4.13
BR-03-06-09	162	2.479	-8.8	31.6	0.2548	80.5000	0.2185	0.322	4.19
BR-04-06-09	216	2.291	-8.2	35.4	0.1914	107.4999	0.1878	0.281	4.04
BR-05-05-05	125	1.981	-10	40.5	0.0000	61.9999	0.1974	0.166	-1.58
CLS-5	5	1.521	-38.3	27.5	0.2433	0.5334	0.5821	0.699	7.68
CLS-10	10	2.09	-19.6	28.3	0.2153	1.9984	0.4093	0.499	4.31
CLS-20	20	2.616	-14.7	23.6	0.0683	4.0050	0.3833	0.351	-1.22
CLS-40	40	3.318	-11.8	18.3	0.0881	8.3540	0.3920	0.290	-3.08
CLS-80	80	3.414	-9.6	19.7	0.0922	15.8311	0.3285	0.256	-2.12
CLS-160	160	3.844	-8	18	0.0412	32.2759	0.3084	0.203	-2.74
$\Delta\%_{global}$			35.8	24.0			$\Delta\%_{global}$		4.3

^aLos valores de referencia REF están normalizados a la esfera de igual volumen (ver texto).

III.3. Propiedades de partículas cilíndricas

En un trabajo anterior de Ortega y García de la Torre [52], fueron desarrolladas una serie de ecuaciones para el cálculo directo de propiedades en disolución para cilindros cortos (cuya relación longitud/diámetro p es menor que 20, es decir $p < 20$) en régimen diluido, incluyendo partículas discoidales ($p < 1$). A continuación, vamos a hallar unas nuevas ecuaciones para cilindros largos que nos permitan extender el rango de p .

En primer lugar, empleando las técnicas de modelado de esferas para el cálculo hidrodinámico, hemos comparado las propiedades de cilindros con longitud L y diámetro d (con $L/p \leq 20$) con cadenas de N esferas tangentes, con distancias centro-centro y diámetro b , y con una longitud total de la cadena $L = Nb$. Se ha relacionado el diámetro del cilindro con el de las esferas, de tal manera que el volumen total del mismo sea igual al volumen total de la cadena, es decir, la suma del volumen de todas las esferas que componen la cadena. Para ello, el diámetro de las esferas debe ser $\sqrt{3/2}$ veces superior al diámetro del cilindro.

$$V_{Cilindro} = \left(\frac{d}{2}\right)^2 \pi L \quad (\text{III.12})$$

$$V_{Cadena} = \frac{4}{3} \left(\frac{b}{2}\right)^3 \pi N \quad (\text{III.13})$$

$$\text{si } b = d\sqrt{3/2} \quad , \quad V_{Cadena} = V_{Cilindro} \quad (\text{III.14})$$

Escogiendo un valor arbitrario de d , b se determina mediante esta relación, y la longitud de la correspondiente partícula cilíndrica es $L = Nb$. Las propiedades del cilindro de longitud L y una relación axial $p = L/d$, son determinadas a partir de las ecuaciones de Ortega y García de la Torre [52], y las de la cadena de esferas empleando el programa HYDRO++, que incluye el mejor tratamiento posible en la obtención del coeficiente de difusión D_t y la viscosidad intrínseca $[\eta]$ [23]. Para los cálculos de los coeficientes rotacionales, se empleó la técnica de la sustitución cúbica hasta el límite que permite HYDRO++ ($N=250$ esferas). El cociente entre las propiedades del cilindro y las propiedades de la cadena de esferas equivalente, se muestran en la Tabla III.4. Estos cocientes son independientes de la elección de d .

En la Tabla III.4 se han obtenido datos hasta $N = 20$, que corresponde a un $p \approx 25$. Puede observarse que los cocientes están muy próximos a la unidad, por lo que podemos

N	p	R_G	D_t	$[\eta]$
1	1.225	1.054	0.986	1.132
2	2.449	1.021	0.972	1.078
3	3.674	1.010	0.986	1.057
4	4.899	1.006	0.989	1.028
5	6.124	1.004	0.991	1.027
7	8.573	1.002	0.994	1.041
10	12.247	1.001	0.999	1.045
15	18.371	1.000	1.010	0.999
20	24.495	0.999	1.003	0.998

Tabla III.4: Relación de las propiedades en disolución del cilindro respecto a las de la cadena de esferas equivalente, variando la relación axial $p = L/d$.

concluir que las propiedades en disolución de la partícula cilíndrica y su equivalente cadena de esfera son prácticamente las mismas. Esta útil consecuencia es fruto de dos elecciones apropiadas en el procedimiento de modelización: (a) la equivalencia en volumen de ambos modelos geométricos, y (b) el adecuado tratamiento hidrodinámico del modelo de esferas, gracias a los últimos avances en la metodología hidrodinámica [23].

En muchos casos, las teorías básicas predicen unas fórmulas generales para las propiedades de cilindros rígidos. El coeficiente de difusión se puede expresar [53]:

$$D_t = \frac{k_B T (\ln p + C_t)}{Q_t \pi \eta_0 L} \quad (\text{III.15})$$

donde η_0 es la viscosidad del disolvente, T la temperatura en grados Kelvin, y k_B es la constante de Boltzmann. Las más rigurosas teorías [41, 54, 55] toman $Q_t = 3$, mientras C_t , que es el denominado término de efecto final, desaparece en cilindros infinitamente largos, cuando $p \rightarrow \infty$.

De manera similar, la viscosidad intrínseca puede formularse como:

$$[\eta] = \frac{Q_\eta \pi N_A L^3}{M (\ln p + C_\eta)} \quad (\text{III.16})$$

donde M es el peso molecular y N_A el número de Avogadro. De nuevo, existe un valor exacto para $Q_\eta = 2/45$ [42, 56, 57], y C_η es otra corrección final.

Las ecuaciones III.15 y III.16 pueden ser escritas de la siguiente manera:

$$\frac{\pi\eta_0LD_t}{k_B T} = \frac{1}{Q_t} \ln p + \frac{C_t}{Q_t} \quad (\text{III.17})$$

$$\frac{\pi N_A L^3}{M[\eta]} = \frac{1}{Q_\eta} \ln p + \frac{C_\eta}{Q_\eta} \quad (\text{III.18})$$

Si representamos el término a la izquierda de las ecuaciones III.17 y III.18 frente a $\ln p$, deberíamos obtener rectas cuya pendiente es igual a $1/3 = 0.333$ y $45/2 = 22.5$ respectivamente. Cuando los resultados de HYDRO++ para las cadenas de esferas con N grande son representados de esta forma (Figura III.7), obtenemos pendientes de 0.3352 (a) y 23.54 (b), muy concordantes con la teoría. Adoptando los valores confirmados de $Q_t = 3$ y $Q_\eta = 2/45$, los términos de efecto final C_t y C_η pueden obtenerse de los resultados de HYDRO++:

$$C_t = \frac{3\pi\eta_0LD_t}{k_B T} - \ln p \quad (\text{III.19})$$

$$C_\eta = \frac{2\pi N_A L^3}{45M[\eta]} - \ln p \quad (\text{III.20})$$

C_t y C_η deben tener un valor asintótico cuando $p \rightarrow \infty$, pero muestran una ligera dependencia con p que puede ser descrita como una dependencia lineal con $1/p$. Ajustes mediante mínimos cuadrados a una función polinómica permiten obtener los resultados finales:

$$C_t = 43.5795/p^2 - 1.9896/p + 0.3249 \quad (\text{III.21})$$

$$C_\eta = 91.1008/p^2 - 6.7930/p - 0.8251 \quad (\text{III.22})$$

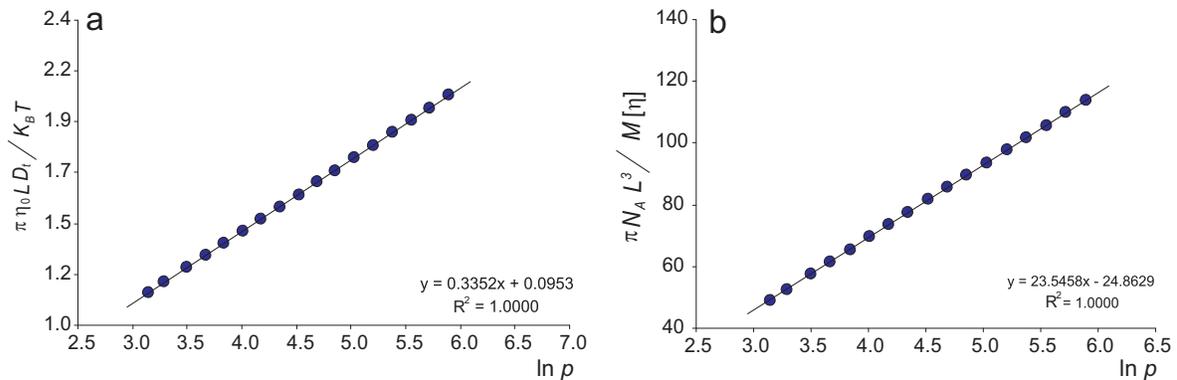


Figura III.7: Representación de las ecuaciones (parte izquierda) III.17(a) y III.18 (b) frente a $\ln p$.

III.3. PROPIEDADES DE PARTÍCULAS CILÍNDRICAS

Cuando $p \rightarrow \infty$, $C_t = 0.325$, valor muy próximo al obtenido años atrás por Tirado y García de la Torre [8, 58] para “verdaderos” cilindros (no cadenas de esferas), $C_t = 0.312$.

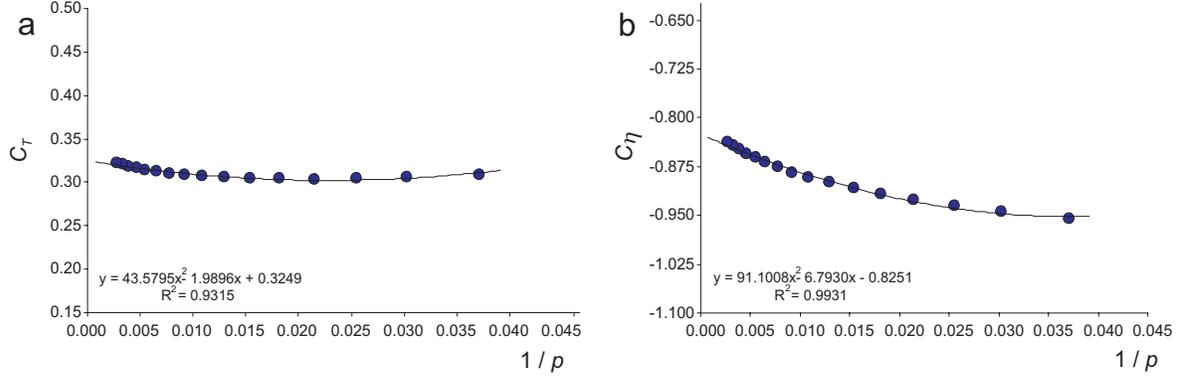


Figura III.8: Ajuste por mínimos cuadrados a a una función polinómica, obteniéndose las ecuaciones III.21 y III.22.

La conclusión de esta comparación es que, empleando HYDR0++, y el modelo de varilla recta de esferas con volumen corregido, se obtienen prácticamente los mismos resultados que para la partícula cilíndrica rígida equivalente, en un amplio rango de relaciones axiales (desde $p = 2$ hasta el límite $p \rightarrow \infty$). Asumimos que esta equivalencia debe conservarse en el caso de cilindros curvados, por lo que concluimos que nuestro modelo discreto de esferas tangentes con volumen corregido es una representación válida para el cilindro vermiforme. La expresión general del coeficiente rotacional de cilindros es [42, 8, 9]:

$$D_r^\perp = Q_r^\perp \frac{k_B T (\ln p + C_r^\perp)}{\pi \eta_0 L^3} \quad (\text{III.23})$$

y la teoría más rigurosa [59] predice que $Q_r = 3$, mientras el término C_r^\perp es función de la relación axial p , como en el caso de C_t , debe tener un valor asintótico para cilindros muy largos. Si se analizan los resultados de D_r^\perp de la cadena de esferas bajo esta forma:

$$\frac{\pi \eta_0 L^3 D_r^\perp}{k_B T} = Q_r^\perp \ln p + Q_r^\perp C_r^\perp \quad (\text{III.24})$$

de nuevo, una representación del término de la izquierda de la ecuación III.24 frente a $\ln p$ (Figura III.9a) debe presentar una pendiente igual a Q_r^\perp . El valor obtenido, $Q_r^\perp = 2.98$, esta muy próximo al de la predicción teórica. Respecto al término de efecto final C_r^\perp , puede obtenerse a partir de:

$$C_r^\perp = \frac{\pi \eta_0 L^3 D_r^\perp}{3 k_B T} - \ln p \quad (\text{III.25})$$

D_t	$k_B T(\ln p + C_t)/3\pi\eta_0 L$	$C_t = 43.5795/p^2 - 1.9896/p + 0.3249$
$[\eta]$	$2\pi N_A L^3/45M(\ln p + C_\eta)$	$C_\eta = 91.1008/p^2 - 6.7930/p - 0.8251$
D_r^\perp	$3k_B T(\ln p + C_r^\perp)/\pi\eta_0 L^3$	$C_r^\perp = 27.4604/p^2 - 0.9467/p - 0.6126$
D_r^\parallel	$k_B T p^2(1 + K_r)/\pi\eta_0 L^3$	$K_r = -1.4/p$

Tabla III.5: Resumen de las ecuaciones obtenidas para el cálculo de las propiedades en disolución de cilindros largos y rígidos.

Para cada valor de D_r^\perp , al cual le corresponde un valor de p , C_r^\perp es obtenido a partir de la ecuación III.25. Encontramos que los valores presentan una ligera dependencia con p , y como se muestran en la Figura III.9b, puede ser representado como:

$$C_r^\perp = 27.4604/p^2 - 0.9467/p - 0.6126 \quad (\text{III.26})$$

Como ocurría con C_t , cuando $p \rightarrow \infty$ obtenemos un $C_r^\perp = 0.613$, valor muy próximo por el obtenido, también para cilindros "verdaderos", de nuevo, por Tirado y García de la Torre [58, 8] ($C_r^\perp = 0.662$).

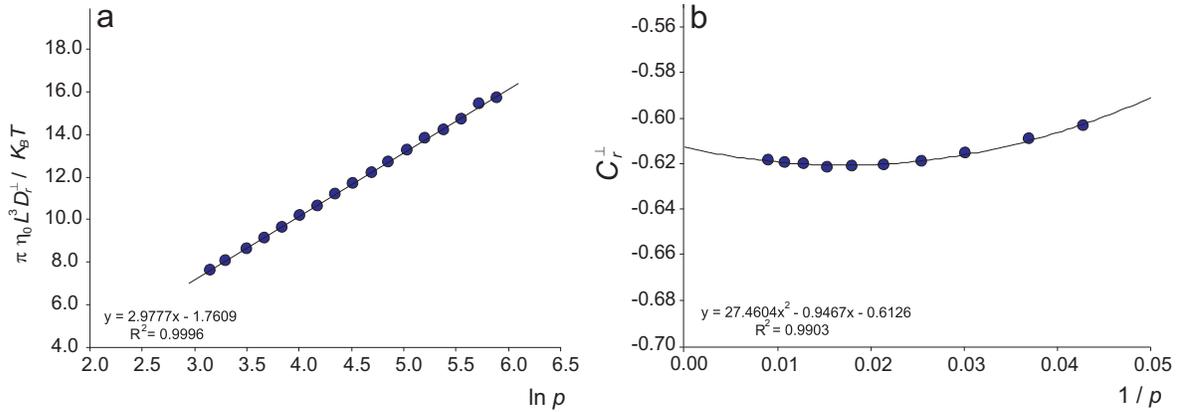


Figura III.9: Ajuste por mínimos cuadrados a una función polinómica, obteniéndose las ecuaciones III.25 y III.26.

Respecto al coeficiente paralelo, una forma útil de expresar su dependencia frente a las dimensiones es la siguiente:

$$D_r^\parallel = Q_r^\parallel \frac{k_B T p^2}{\pi\eta_0 L^3} (1 + K_r) \quad (\text{III.27})$$

donde, como en los casos anteriores, Q_r es un factor numérico y el término K_r contiene la dependencia con p . El problema con D_r^\parallel es que está muy influido por la descripción de la

III.3. PROPIEDADES DE PARTÍCULAS CILÍNDRICAS

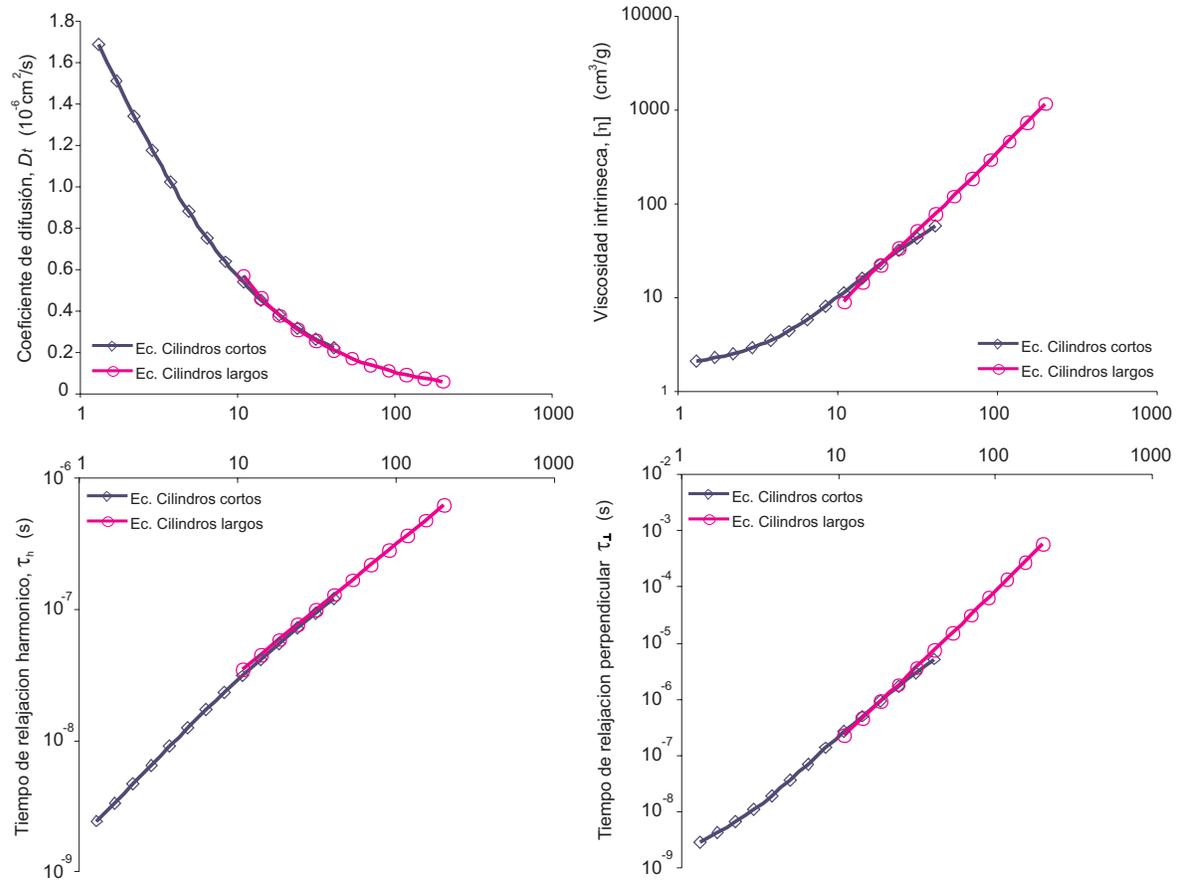


Figura III.10: Representación de las ecuaciones de Ortega-García de la Torre [52], junto a las obtenidas anteriormente, cubriendo todo el rango de p . Nótese la intersección en $p = 20$.

sección transversal (porque el material que compone la partícula está mucho más próximo a este eje que al perpendicular D_r^\perp). Por lo tanto, no es apropiado reemplazar la cadena de esferas tangentes por una cadena de cubos, de acuerdo a los resultados anormales obtenidos (datos no mostrados). Por ello, en este caso hemos optado por el modelo de concha para representar la partícula cilíndrica, empleando HYDROSUB [25] para el cálculo de D_r^\parallel . En los cálculos realizados hasta $p = 70$, se obtienen valores de $\frac{D_r^\parallel \pi \eta_0 L^3}{k_B T p^2} = Q_r^\parallel (1 + K_r)$ en el rango 0.87 - 0.97, con una dependencia con p no muy bien definida (datos no mostrados). Por ejemplo, obtenemos que $\frac{D_r^\parallel \pi \eta_0 L^3}{k_B T p^2} = 0.97$ para $p = 20$, que fue tomado por Ortega y Garcia de la Torre [52] como límite en los cálculos de HYDROSUB para cilindros. Por ello, es razonable tomar $Q_r^\parallel = 1$ y dejar la dependencia con p al término K_r , que toma valores de $K_r = -0.07$ para $p = 20$, para poder reproducir los resultados mencionados anteriormente, con el límite $K_r = 0$ cuando $p \rightarrow \infty$. Asumiendo una dependencia con $1/p$, expresamos K_r como:

$$K_r = -1.4/p \quad (\text{III.28})$$

En la Figura III.10 se representan las ecuaciones para cilindros cortos [52], junto a las obtenidas para cilindros largos, y la intersección entre ambas.

En resumen, las propiedades en disolución de cilindros largos y rígidos pueden obtenerse a partir de las ecuaciones de la Tabla siguiente (Tabla III.5). Con ello, como indicábamos, se complementa las teorías anteriores para cilindros cortos.

III.4. Potenciales torsionales para cadenas flexibles

Otro avance metodológico en esta Tesis, que ha sido necesaria para la simulación de de cadenas polipetídicas flexibles que se presentará más adelante, se refiere a los potenciales torsionales y a su incorporación en los programas MONTEHYDRO y BROWFLEX. Los potenciales de torsión suelen expresarse como funciones trigonométricas, y son fundamentales para determinar con cierto detalle la características conformacionales de macromoléculas flexibles. Existen diversas expresiones para este tipo de potencial, que implican un número de parámetros diferente, dependiendo de la estructura química de la macromolécula y del modelo empleado,

Uno de los más utilizados, concretamente en el caso de una torsión alrededor de enlaces entre átomos de carbono con geometría sp^3-sp^3 , es el propuesto por Ryckaert y Bellemans [60]:

$$V_{\phi}(\phi_i) = \sum_{k=0}^5 c_k (\cos\phi_i)^k \quad i = 1, \dots, N - 3 \quad (\text{III.29})$$

donde ϕ_i es el valor instantáneo del ángulo diedro, y c_k son constantes. La denominación que emplearemos de aquí en adelante para las conformaciones de ángulos diedros son *trans* para valores de $\phi = 0^\circ$, *gauche +* para valores de $\phi = 120^\circ$, *gauche -* para valores de $\phi = 240^\circ$ y *cis* para valores de $\phi = 180^\circ$.

Otra expresión muy empleada para representar este tipo de potencial [61] es la siguiente:

$$V_{\phi}(\phi_i) = K_{\phi}[1 + \cos(3\phi_i)] \quad i = 1, \dots, N - 3 \quad (\text{III.30})$$

donde K_{ϕ} es una constante. Para esta torsión alrededor de enlaces $sp^3 - sp^3$ también se puede emplear [62]:

$$V_{\phi}(\phi_i) = \frac{K_{\phi}}{2}[x(1 - \cos\phi_i) + (1 + x)(1 - \cos(3\phi_i))] \quad i = 1, \dots, N - 3 \quad (\text{III.31})$$

Si la rotación se da alrededor de un enlace C-C $sp^3 - sp^2$ (impedida), podemos utilizar la expresión del potencial propuesto por Rey et al [63]:

$$V_{\phi}(\phi_i) = c_1(1 + \cos\phi_i) + c_2(1 - \cos(2\phi_i)) + c_3(1 + \cos(3\phi_i)) \quad i = 1, \dots, N - 3 \quad (\text{III.32})$$

En el caso de una rotación impedida en torno a un enlace C-C $sp^2 - sp^2$, emplearemos el siguiente potencial del mismo autor [63]:

$$V_{\phi}(\phi_i) = c_1[(\cos\phi_i) - (\cos\phi_0)] \quad i = 1, \dots, N - 3 \quad (\text{III.33})$$

donde ϕ_0 es el ángulo diedro de equilibrio.

A continuación, uno de los potenciales de torsión más empleados para modelos ‘coarse grained’ de proteínas, perteneciente al modelo *Gō-like* [64, 65], cuya expresión coincide (y es aplicable) con otro modelo propuesto por Tozzini *et al* [66], para simular una distribución “genérica” donde todas las estructuras secundarias son representadas. En ambos casos, se trata de un potencial definido entre cuatro átomos C α consecutivos de la proteína:

$$V_\phi(\phi_i) = c_0[1 - \cos(\phi_i - \phi_{0i}^{(1)})] + c_1[1 - \cos 3(\phi_i - \phi_{0i}^{(3)})] \quad i = 1, \dots, N - 3 \quad (\text{III.34})$$

donde c_0 y c_1 son constantes, ϕ_i es el ángulo de torsión a un instante dado y ϕ_0 es el ángulo de torsión que presenta en la estructura nativa de la proteína, en el modelo *Gō-like*. En el caso del potencial genérico de Tozzini *et al*, $\phi_0 = 0^\circ$.

Autor	Ec.	c_0	c_1	c_2	c_3	c_4	c_5	Ref.
Ryckaert <i>et al</i>	III.29	1.941	3.730	-1.071	-3.703	2.141	-3.034	[63]
Lamm <i>et al</i>	III.30	-	1.0	-	-	-	-	[61]
Winkler <i>et al</i>	III.31	-	4.1	-	-	-	-	[62]
Rey <i>et al</i>	III.32	-	-0.853	-0.111	-2.320	-	-	[63]
Rey <i>et al</i>	III.33	-	1240	-	-	-	-	[63]
Clementi <i>et al</i>	III.34	1.0	0.5	-	-	-	-	[65]
Tozzini <i>et al</i>	III.34	0.5	1.0	-	-	-	-	[66]

Tabla III.6: Parámetros de los potenciales de torsión para simulaciones de moléculas poliméricas, donde se definen los potenciales diedros entre cuatro carbonos. También se incluyen los parámetros de los modelos simplificados de proteínas ($\text{Kcal} \cdot \text{mol}^{-1}$).

Simulación de la dinámica de una cadena de átomos de Carbono

Antes de comenzar la simulación dinámica de una macromolécula compleja, se debe comprobar el funcionamiento de MONTEHYDRO y BROWFLEX en lo que respecta a los potenciales de torsión, de reciente incorporación en el programa. Para ello, lanzaremos una serie de simulaciones cortas, de modelos sencillos, pero lo suficientemente largas para obtener resultados estadísticamente aceptables.

La prueba consiste en reproducir las distribuciones de frecuencia de ángulos diedros, y evaluar su correspondencia con la función de potencial, que mediante la inversión de Boltzmann podemos transformar a distribución de probabilidad, según la ecuación $V(q) = -k_B T \ln(P(q))$, o $P(q) = \exp(-V(q)/k_B T)$, siendo q una coordenada interna, como ángulo o posición, k_B la constante de Boltzmann, T la temperatura absoluta, $V(q)$ y $P(q)$ son el potencial de energía y la distribución de probabilidad/frecuencia en el punto q respectivamente. Previa normalización, $P(q)$ deberá coincidir con la distribución de frecuencias obtenida de la simulación. En lo que aquí respecta, q es el ángulo diedro ϕ .

Respecto a los detalles de la simulación, se construyó un modelo de 16 carbonos, donde cada esfera representa un grupo de átomos (grupo metilo). Nuestro modelo se compondrá de 16 elementos, unidos por 15 enlaces. Definimos 14 ángulos entre elementos contiguos y 13 ángulos diedros, formados por 4 enlaces consecutivos. Se analizaron todos los ángulos diedros presentes en cada conformación, y se emplearon los parámetros que aparecen en la Tabla III.6.

Se emplearon las mismas condiciones que en la simulación anterior para los cuatro potenciales de ángulo diedro restantes, dando el resultado esperado en los cuatro casos. Los resultados se muestran en la figura III.12.

Simulación de la dinámica conformacional de un péptido

En el primer caso, utilizamos el potencial III.34 aplicado al modelo *Gō-like*. La prueba consiste en construir un péptido de 32 aminoácidos, representado por una cadena virtual constituida por los $C\alpha$. Empleamos un potencial armónico de enlace, centrado en 3.8 Å, y otro potencial angular armónico centrado en 112.5° (ver capítulo V). Empleamos un ángulo diedro de equilibrio correspondiente a una región extendida ($\phi_0 = 180^\circ$) en todos los ángulos de torsión de la molécula. En el segundo caso, aplicamos el potencial III.34 al modelo de potencial genérico de Tozzini. Empleando el mismo sistema que en el caso anterior, únicamente variando los parámetros y el ángulo de equilibrio ($\phi_0 = 0^\circ$). En ambos casos, la conformación inicial es extendida y al azar, y se simula mediante dinámica browniana durante un tiempo de 2 μs , precedido de un periodo de ‘calentamiento’ de 20 ns. Empleamos los parámetros recogidos en la Tabla III.6. Para hacer las distribuciones, se analizaron todos los diedros presentes. Los resultados se muestran en la figura III.13.

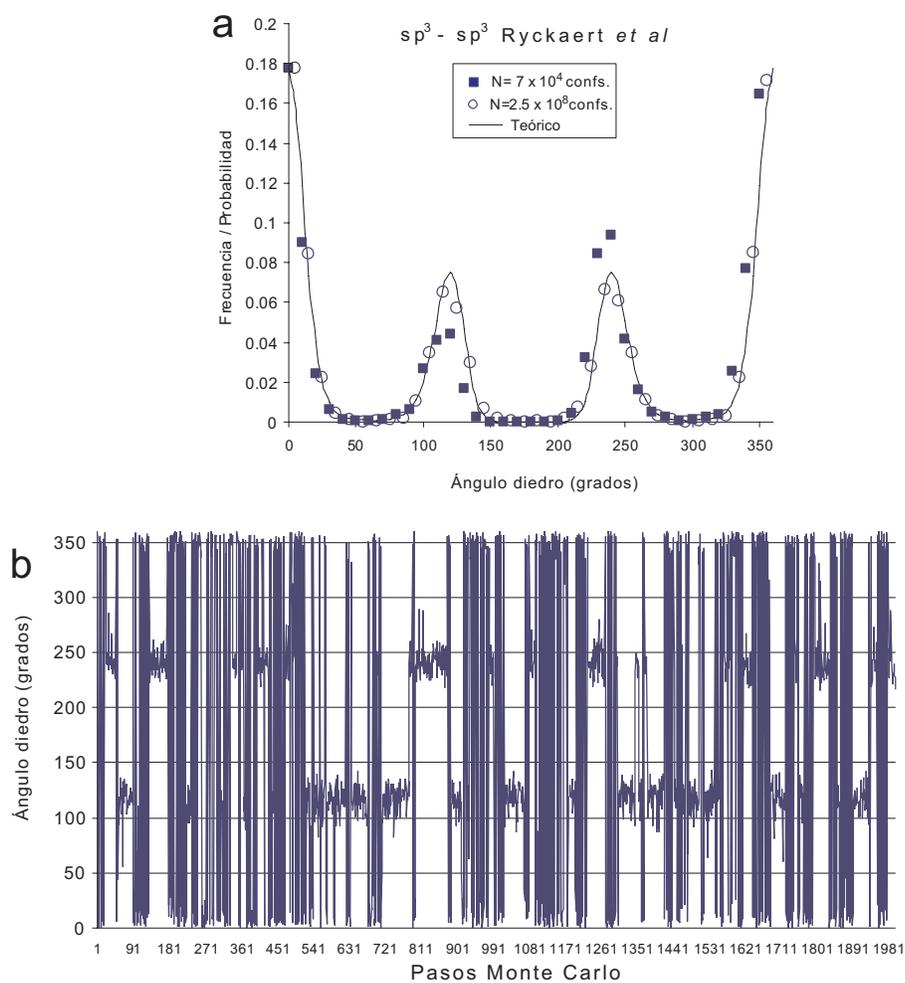


Figura III.11: (a) Comparación de la distribución de frecuencias obtenida de dos simulaciones con diferente número de conformaciones, respecto a la curva analítica de la función potencial III.29 [60]. Nótese que es necesario un número elevado de conformaciones (2.5×10^8 conf.) para poder reproducir la distribución esperada, mientras que la prueba realizada con 7.0×10^4 conformaciones es evidentemente insuficiente. (b) En la parte inferior podemos observar un fragmento de la simulación, donde se observan las transiciones del ángulo de torsión, concluyendo que el número de pasos y conformaciones de Monte Carlo empleado es suficientemente elevado para observar numerosas transiciones.

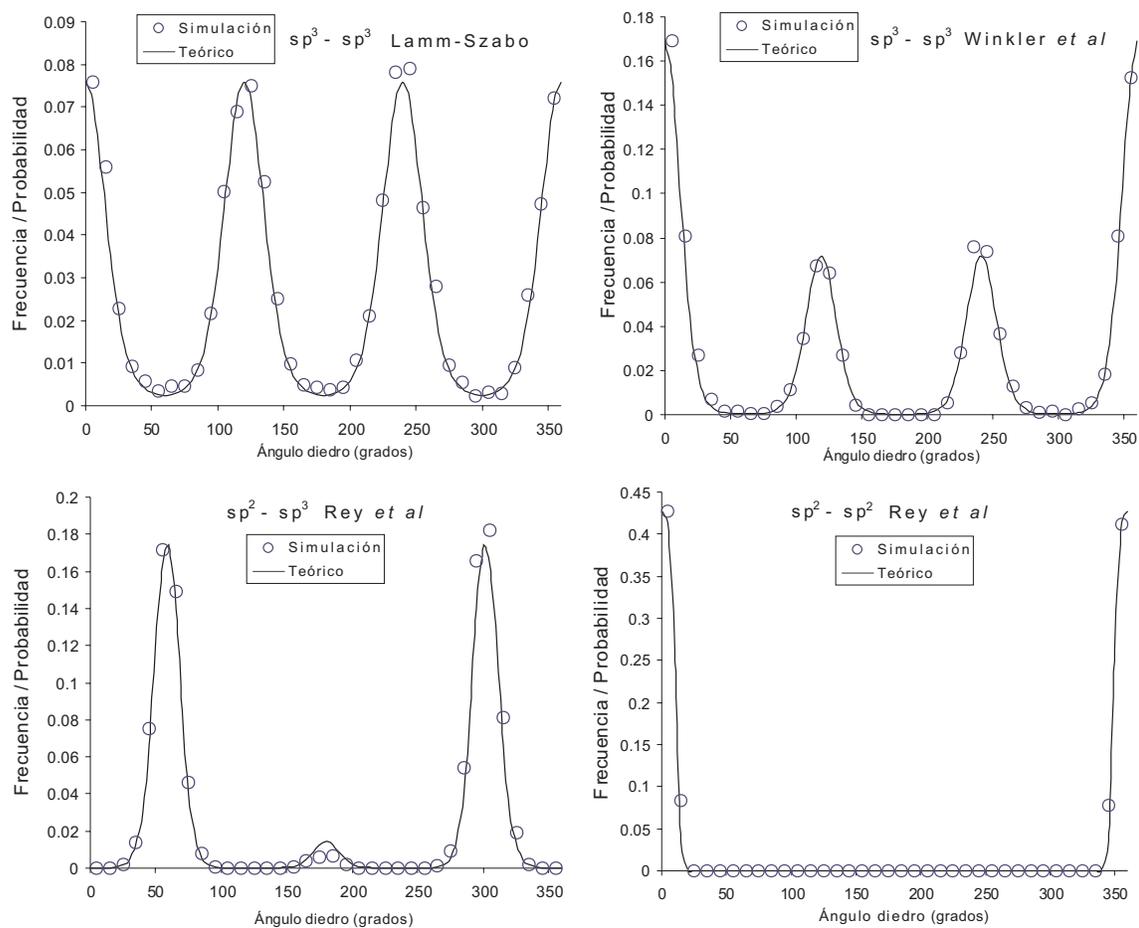


Figura III.12: Comparación de la distribución de frecuencias obtenida de dos simulaciones con diferente número de conformaciones, respecto a la curva analítica de las funciones potencial III.30, III.31, III.32 y III.33.

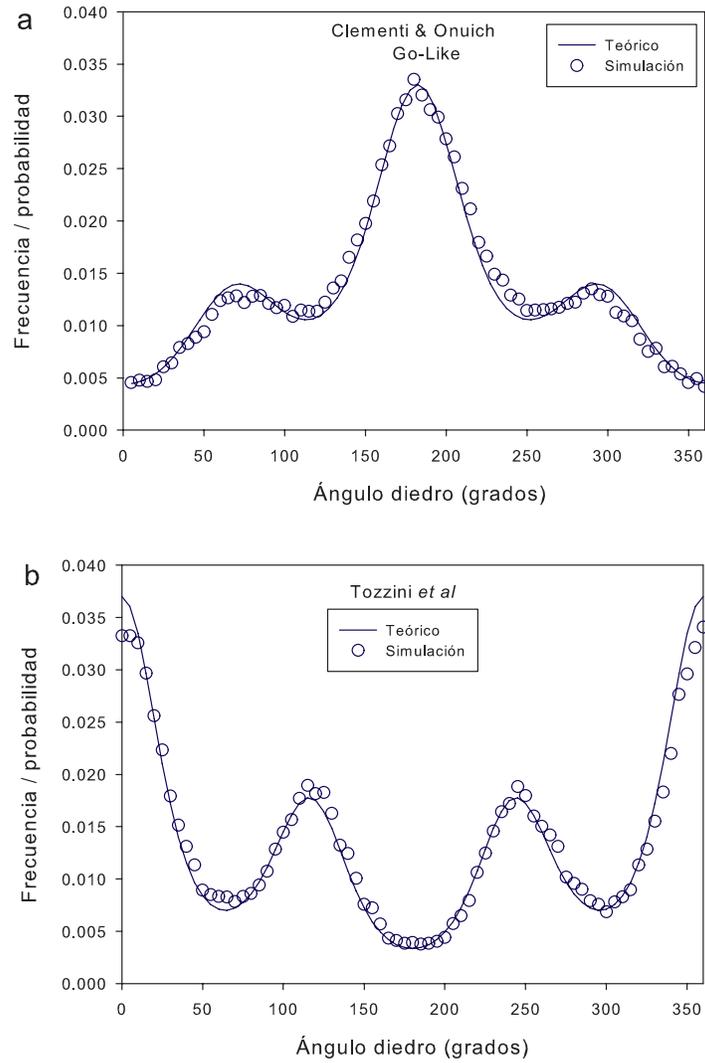


Figura III.13: Comparación de la distribución de frecuencias obtenida respecto a la curva analítica de la función potencial III.34. (a) Modelo *Gō-like* [65], (b) modelo de Tozzini *et al* [66].

III.5. HYDROFIT: de las propiedades a la estructura

Introducción

Actualmente, impulsado por los avances de la ciencia y de la computación, el modelado por esferas se ha desarrollado en todos los posibles niveles. A partir del primer programa disponible para casos generales, HYDRO [22] (recientemente reformado ya ampliado en HYDRO++ [23]), es ahora posible calcular simultáneamente todas las propiedades hidrodinámicas y de dispersión de luz a partir de modelos de esferas de nivel atómico, obteniéndose estos modelos a partir de archivos de coordenadas atómicas, como en nuestros programas HYDROPRO e HYDRONMR, u otros similares desarrollados por otros autores [13, 67, 68, 69, 70, 71]. También es posible utilizar, con el mismo propósito datos de microscopía electrónica (mapas de densidad electrónica de alta resolución) mediante el programa HYDROMIC [67].

En el caso de grandes proteínas y complejos macromoleculares, no es posible el cálculo a nivel atómico de las propiedades en disolución, debido a la ausencia de datos cristalográficos o de NMR, o, simplemente, la estructura estática a nivel atómico, tan detallado para estructuras tan grandes, contribuye muy poco en el comportamiento hidrodinámico. Para estos casos, en los que la estructura está compuestas por varias subunidades, como los anticuerpos, que denominamos genéricamente estructuras multiméricas, son el tamaño y forma global de las subunidades, así como la disposición espacial, los aspectos que determinan las propiedades en disolución. Cuando las subunidades se representan como elipsoides o cilindros, el modelado de concha de esferas en este nivel se realiza fácilmente con nuestro programa HYDROSUB [25]. Estructuras con otras formas geométricas pueden ser tratadas con nuestro programa HYDROPIX [72]. Cabe mencionar el cálculo de otras propiedades no hidrodinámicas, como el tensor de giro y el tensor de orientación para NMR, que realiza el programa ORIEL [73].

En resumen, de una u otra manera, estos métodos posibilitan el cálculo de propiedades a partir de estructuras. Sin embargo – como ya se apuntaba en el capítulo I – el paradigma central de la biología estructural es el de la determinación de estructuras a partir de propiedades, y ya hemos indicado como las propiedades hidrodinámicas tienen un contenido de información escaso para una determinación estructural detallada, aunque pueden usarse para la elucidación de algunos aspectos, refinamiento o validación de estructuras, etc. Esta posibilidad fue anticipada en un trabajo previo (A. Ortega, Tesis

Doctoral [74]), centrado en modelos sencillos rígidos (elipsoides, cilindros) o flexibles (cadenas vermiformes), y materializada en los programas **Single-HYDFIT** y **Multi-HYDFIT** [75]. En estos modelos sencillos, la especificación estructural se reduce a la asignación unos pocos parámetros estructurales, como los ejes del elipsoide, la longitud y diámetro del cilindro, o los tres parámetros del modelo vermiforme (que se trata en el capítulo IV).

Para analizar (ajustar) conjuntamente datos de propiedades tan diversas como coeficientes de sedimentación o difusión, radio de giro, viscosidad intrínseca, covolumen, etc, se propuso utilizar [75], en lugar de estas, sus correspondientes radios equivalentes, dado que estos dependen de las características estructurales de una manera más uniforme que las propiedades. El grado de concordancia entre un conjunto de datos experimentales, y los calculados, puede ser cuantificado en términos de radios equivalentes por medio de la expresión,

$$\Delta^2 = \left(\sum_X w_X \right)^{-1} \sum_X w_X \left[\frac{a_X(cal) - a_X(exp)}{a_X(exp)} \right]^2 \quad (\text{III.35})$$

donde \sum_X indica una suma sobre todas las propiedades listadas en la Tabla II.1, cuyos datos fuesen conocidos. Nótese que Δ^2 es la desviación cuadrática media relativa entre los radios equivalentes calculados y los experimentales. w_X es un peso estadístico o factor importancia que puede ser asignado a cada propiedad, de manera que Δ^2 es un promedio ponderado. Estos pesos pueden utilizarse si, por alguna circunstancia bien definida (por ejemplo, en términos de los posibles incertidumbres en los datos experimentales) se desea ponderar en mayor o menor grado las diversas propiedades. Pero salvo en tales casos, gracias al uso de radios equivalentes, es aceptable no considerarlos, es decir, tomar $w_X = 1$ para cualquier propiedad.

Además, es de notar que se emplean desviaciones relativas para que la estimación no dependa del tamaño de la partícula, sino solamente de su forma. $100\sqrt{\Delta^2}$ es el porcentaje de desviación del radio equivalente calculado respecto al experimental, y sirve como medida global de la bondad del ajuste entre la estructura supuesta y el conjunto de datos experimentales.

Otra posibilidad para evaluar la bondad del ajuste de la estructura asumida es mediante el empleo de las relaciones de radios. Se define la siguiente función:

$$\nabla^2 = \left(\sum_{XY} w_{XY} \right)^{-1} \sum_{XY} w_{XY} [XY(cal) - XY(exp)]^2 \quad (\text{III.36})$$

donde ∇^2 es la diferencia cuadrática media entre las relaciones de propiedades calculadas y las relaciones de las experimentales, y w_{XY} son pesos estadísticos a los que se aplican los criterios indicados para los XY . De nuevo, $100\sqrt{\nabla^2}$ es el porcentaje de desviación de la relación calculada respecto a la experimental.

Así, en el mencionado trabajo reciente [75], se propuso emplear Δ o ∇ como funciones objetivo, que deben ser minimizadas en la búsqueda conformacional de la estructura que mejor ajuste al conjunto de las propiedades experimentales, implementadas en los programas `Single-HYDFIT` y `Multi-HYDFIT` para los modelos macromoleculares sencillos: elipsoides, cilindros y cadenas vermiformes. Recordamos que `Multi-HYDFIT` está diseñado para tratar el caso específico de una serie homóloga con múltiples datos, de un conjunto de muestras de idéntica estructura general, aunque de tamaños variables, situación que será de aplicación a lo que trataremos en el capítulo IV, sobre propiedades de cadenas vermiformes.

En esta Tesis hemos pretendido generalizar el concepto de ajuste global a la búsqueda estructural no ya en el caso de modelos tan simplificados, sino en el caso de partículas de forma arbitraria, cuyas propiedades en disolución son calculadas, como antes describíamos, a partir programas de los programas basados en modelos de esferas, tales como `HYDRO++`, `HYDROPRO`, `HYDROSUB`, entre otros. Ciertamente, no podemos pretender que, con la información contenida en unas cuantas propiedades, se pueda construir de forma *ab initio* el modelo molecular, pero sí, sobre la base de un modelo genérico, ajustar las características del mismo. La estrategia a seguir consiste en generar un gran número de modelos de esferas, variando de uno a otro los parámetros o características que están siendo objeto de estudio, y efectuando, para cada uno de esos muchos modelos, los cálculos con el programa que proceda. Los programas de la serie `HYDROxxx` admiten la ejecución consecutiva de un número indefinido de cálculos para sucesivos modelos. No obstante, la construcción de los ficheros de datos – sobre modelos comunes en los que simplemente varían algunos datos – para tales multi-ejecuciones puede resultar un tanto tediosa de manera manual. Con este objetivo hemos construido unos programas auxiliares, tales como `MULTIHYDRO` o `MULTISUB`, aunque el usuario podría escribir sus propios programas con este propósito. Por otra parte, los programas de la familia `HYDROxxx` han sido modificados para generar

un archivo de salida, denominado `hydfitdat.txt`, que recopila los resultados de todas las múltiples estructuras procesadas en una ejecución.

Entonces, el nuevo programa HYDROFIT, desarrollado como parte de esta Tesis, se encarga del tratamiento de este conjunto de múltiples resultados. Este nuevo programa puede considerarse como una secuela de `Single-HYDFIT` y `Multi-HYDFIT`. Su propósito es, en primer lugar, captar un archivo que contiene los valores de las propiedades experimentales, y por otro lado el procedente de las ejecuciones multi-caso `hydfitdat.txt`, siendo así capaz de determinar para cada estructura el valor de Δ correspondiente, (ec. III.35), y encontrar aquella que de manera óptima ajusta el conjunto de datos experimentales.

Caso de una estructura de subunidades. Determinación de la estructura de ciertos anticuerpos

Vamos a presentar una descripción detallada de la estrategia apoyada en HYDROFIT, para el caso de grandes proteínas y complejos macromoleculares, para las que no es posible el cálculo a nivel atómico de las propiedades en disolución, debido a la ausencia de datos cristalográficos o de NMR, o, simplemente, la estructura estática a nivel atómico contribuye muy poco en el comportamiento hidrodinámico, como antes indicábamos. Proteínas compuestas por varias subunidades, como los anticuerpos, son ejemplos de este tipo, que denominamos genéricamente estructuras multiméricas. El tamaño y forma global de las subunidades, así como la disposición espacial, son los aspectos que determinan las propiedades en disolución.

Para tales casos, fue diseñado el programa HYDROSUB, [25], que persigue un enfoque a media escala o meso-escalar, que ha cobrado popularidad actualmente, ya que permite ignorar los detalles innecesarios y centrarse en los aspectos esenciales. Las subunidades son representadas por elipsoides o cilindros, poniendo énfasis en sus dimensiones y orientaciones espaciales relativas. La posibilidad de ejecuciones multi-caso, variando alguna característica o parámetro de estos modelos de subunidades ya fue contemplada en una versión previa del programas MULTISUB, [33] que, en conjunción con HYDROSUB ha sido utilizado ya previamente – sin las conveniencias del nuevo programa HYDROFIT – para moléculas de anticuerpo [76, 77, 78, 79, 80, 81, 82].

Un esquema del nuevo procedimiento global se indica en la Figura III.14. En primer lugar, debemos determinar para cada subunidad, sus dimensiones (ejes de revolución de los elipsoides, longitud y diámetro de los cilindros). Esto puede realizarse por dos vías

(a) a partir de los datos experimentales de propiedades en disolución de las subunidades aisladas (b) o bien a partir de estructuras detalladas (atómicas, de microscopía o de SAXS/SANS) conocidas para las subunidades individuales, ya que programas como HYDROPRO o HYDROMIC permiten predecir con fiabilidad tales propiedades. Entonces, el programa ya disponible, **Single-HYDFIT** permite realizar el correspondiente ajuste de los datos a las dimensiones de la subunidad.

Una vez que las dimensiones de las subunidades individuales han sido determinadas, ya podemos construir la maqueta o modelo de la partícula completa, aún dependientes de algún detalle estructural variable. A continuación, una versión adaptada al problema del programa MULTISUB genera las numerosas conformaciones del modelo, entre las que varían los parámetros estructurales que se están investigando, produciendo los numerosos ficheros de datos para una ejecución multi-caso de HYDROSUB. Esta, a su vez, produce el conjunto de resultados, compactados en el fichero `hydfitdat.txt`, que es finalmente procesado, junto a los datos experimentales de la macromolécula global, mediante HYDROFIT para determinar la estructura óptima.

Este esquema pertenece a lo que actualmente se viene denominando “modelado multi-escala”, donde el sistema completo de la partícula es representado por un modelo de baja resolución o de “grano grueso”, constituido por elementos simples, cuyos parámetros derivan generalmente de una simulación de alta resolución o “grano fino” de sus propiedades en disolución.

En lo sucesivo abordamos la determinación estructural de anticuerpos, con algún nuevo dato experimental, y sobre todo como utilización y ejemplo de la estrategia de cálculos multi-modelo y optimización con HYDROFIT.

Las inmunoglobulinas (Ig) o anticuerpos son un grupo de glucoproteínas producidas por las células plasmáticas (último estadio diferenciativo de los linfocitos B). La estructura molecular de las mismas las capacita para desempeñar su principal función biológica: el reconocimiento de antígenos extraños. Así, las Ig son lo suficientemente diversas y heterogéneas para poder reconocer los numerosos antígenos existentes. Las moléculas de anticuerpo están compuestas por 4 cadenas: 2 cadenas pesadas (H) idénticas de 50-77 KDa y 2 cadenas ligeras (L) idénticas de 25 KDa, unidas por enlaces covalentes (puentes disulfuro) y fuerzas no covalentes. Además, cada tipo de cadena presenta una región variable (V), y otra constante (C).

La degradación enzimática de las Ig permite subdividir las en 2 fragmentos biológicamente activos: Fab y Fc. El fragmento Fab (“fragment antigen binding”) o zona de combinación

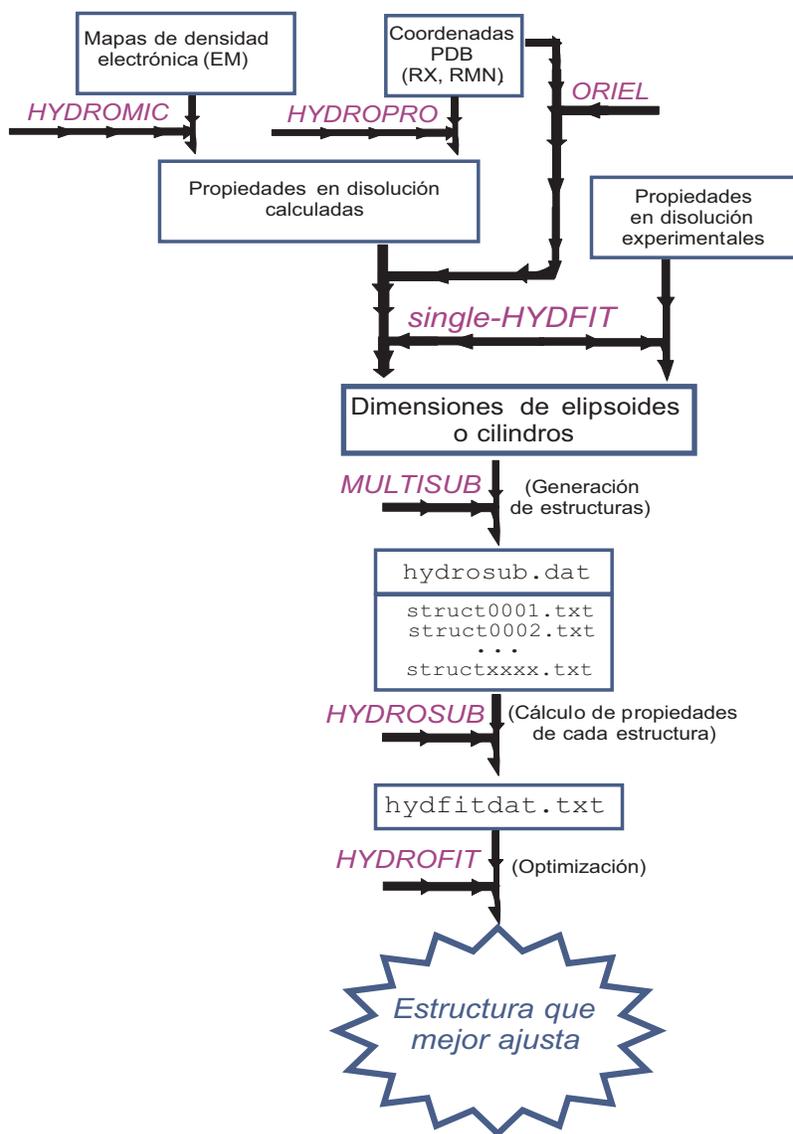


Figura III.14: Esquema del procedimiento para determinar la conformación global de una estructura multimérica. El nombre de los programas empleados en cada paso aparecen en rojo.

con el antígeno, está constituido por una cadena ligera, la región variable y el primer dominio de la región constante de una cadena pesada. El fragmento Fc (fracción cristalizante), formado por los dominios CH₂ y CH₃ unidos por un puente disulfuro, es el responsable de las funciones biológicas de las inmunoglobulinas, tales como activación del complemento, paso a través de la placenta y unión a células del sistema inmune. Ambos fragmentos están conectados mediante una región denominada bisagra, la cual, debido a su composición de aminoácidos (prolina, serina y treonina), dota de gran flexibilidad a la molécula, proporcionando movilidad a los fragmentos Fab y facilitando la combinación con el antígeno.

Diferenciación de la estructura de las formas nativa y mutante de IgG

La elucidación de la estructura global del anticuerpo IgG3 es el ejemplo seleccionado para ilustrar nuestro planteamiento. En una primera etapa, debemos hallar las dimensiones de los modelos hidrodinámicos que van a representar las subunidades Fab y Fc. A continuación, determinar la disposición espacial de estas subunidades en la molécula completa.

Los datos experimentales disponibles para Fab y Fc son escasos; las únicas propiedades en disolución de las que disponemos, para ambos fragmentos, son el coeficiente de sedimentación s , y el radio de giro R_g . Los valores se muestran en la Tabla III.7. Por lo tanto, tenemos una información muy limitada para la determinación de las dimensiones de los elipsoides, los parámetros a y b , que corresponden a los semiejes de dichos elipsoides.

Afortunadamente, disponemos de otra fuente de información, sus estructuras atómicas de alta resolución publicadas: las estructuras cristalinas de las subunidades Fab y Fc (archivos PDB **1BBJ** [83] y **1FC1** [84]). Las propiedades en disolución de estas unidades fueron predichas empleando para ello el programa HYDROPRO con valor estándar de 3.5 Å como radio de hidratación atómica, obteniéndose valores para los coeficientes de difusión y sedimentación, s y D_t , radio de giro, R_g , así como de la viscosidad intrínseca $[\eta]$, el tiempo de relajación más largo, τ_1 , el covolumen u , y la cuerda más larga, L_{max} (o D_{max}). Decidimos realizar el ajuste de los parámetros del modelo hidrodinámico elipsoidal, a partir de toda esta amplia gama de propiedades que, a partir de la estructura atómica, predice HYDROPRO.

Este conjunto de propiedades de las subunidades Fab y Fc, obtenidas de HYDROPRO, son los datos de partida de **Single-HYDFIT**, para encontrar los correspondientes elipsoides

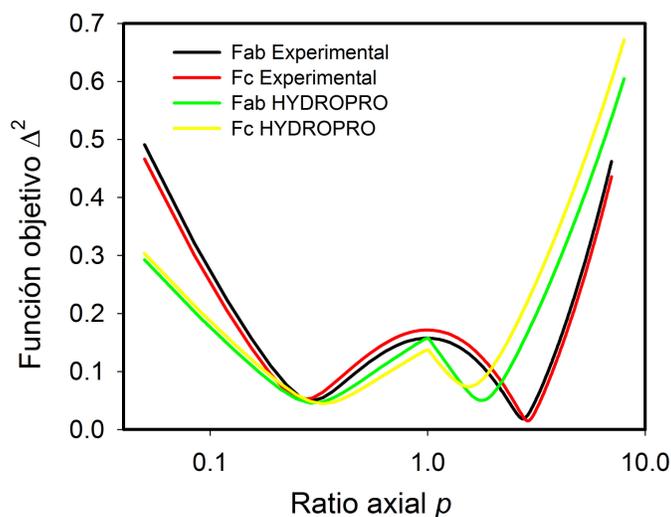


Figura III.15: Variación de la función objetivo ∇ con el parámetro de forma del elipsoide. (Radio axial $p = a/b$).

equivalentes. En primera instancia, empleamos un modo de ajuste en el cual sólo consideramos $p = a/b$ como único parámetro, en una optimización independiente del tamaño, y que sólo depende de la forma de la molécula. Para ello empleamos las relaciones de radios, basada en la optimización de la función objetivo ∇ (ec. III.36). **Single-HYDFIT** halla el mínimo, como variación de ∇ frente al parámetro p .

En la Figura III.15 se muestran los resultados obtenidos, donde se revela la existencia de un mínimo absoluto y otro mínimo secundario. Esta circunstancia está relacionada con la ambigüedad existente a la hora de representar proteínas globulares mediante elipsoides, conocida desde los primeros trabajos en estos modelos [19]: dos elipsoides, uno prolato y otro oblato reproducen ambos, de manera favorable, los datos experimentales. En un trabajo previo, se describió este problema con lisozima como ejemplo [75], y propusimos usar datos complementarios, concretamente con un enfoque puramente geométrico, hallando el elipsoide de revolución que mejor se ajuste a los momentos de inercia de la estructura atómica. Este procedimiento está implementado en la herramienta computacional **ORIEL** [73]. Para la subunidad Fab, los datos obtenidos de **HYDROPRO** revelan la existencia de un mínimo absoluto a $p = 1.8$ (prolato), y un mínimo secundario a $p = 0.3$ (oblato). Por otro lado, **ORIEL**, que por su parte da un resultado muy próximo a $p = 1.8$, nos permite rechazar la forma oblato para esta subunidad. Respecto a la subunidad Fc, de nuevo encontramos dos mínimos, $p = 1.5$ y el secundario en $p = 0.32$. El resultado de **ORIEL**, $p = 0.52$, difiere

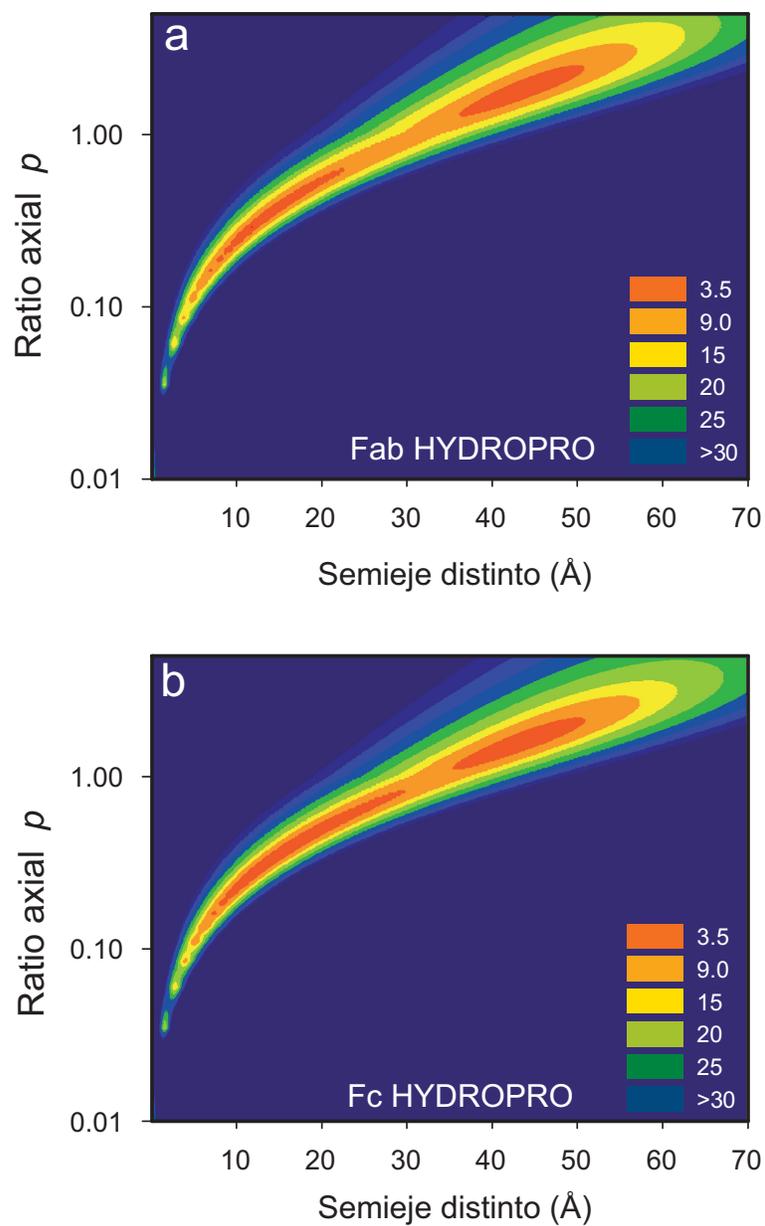


Figura III.16: Mapa de contorno de la función objetivo Δ frente a p y a . Nótese que la escala de p es logarítmica, con el fin de visualizar de forma equitativa la región oblato y prolato. (a)Fab. (b)Fc

Origen de datos	Fab'	Fc
Estructura PDB	1BBJ	1FC1
HYDROPRO : a, b (Å) ; p / (Δ)	43.7 , 24.9 ; 1.80 / 0.043	14.5 , 43.0 ; 0.32 / 0.027
ORIEL : a, b (Å) ; p	42.3 , 23.1 ; 1.83	20.0 , 38.2 ; 0.52
Elipsoide : a, b (Å) ; p / (Δ)	43.7 , 24.9 ; 1.76 / 0.043	17.1 , 41.9 ; 0.41 / 0.035
Exp.: M (g/mol) , \bar{v} (cm ³ /g)	47500 ^[85] , 0.725 ^[76, 77, 85]	51000 ^[77] , 0.732 ^[77]
s (S) : Elipsoide / Exp.	3.58 / 3.6 ^[85] , 3.92 ^[76, 77]	3.63 / 3.85 ^[77]
R_g (Å) : Elipsoide / Exp.	25.1 / 26.9 ^[85] , 28.2 ^[82]	26.9 / 29.3 ^[85]

Tabla III.7: Resultados obtenidos por HYDROPRO y ORIEL para los modelos elipsoidales de Fab y Fc, siendo la entrada “Elipsoide” los valores finalmente empleados para las dimensiones del elipsoide equivalente de cada subunidad, resultado de promediar los valores obtenidos por dichos programas. En las dos últimas líneas se muestra la comparación de las propiedades calculadas del elipsoide equivalente con los valores experimentales obtenidos de Carrasco *et al* [76, 77], Lu *et al* [82] y Byron *et al* [85]

de los obtenidos mediante HYDROPRO. Sin embargo, la estimación de ORIEL nos decanta hacia la forma oblato. Decidimos emplear la forma oblato, promediando los resultados de HYDROPRO y ORIEL, con un p definitivo de 0.42.

A continuación, realizamos otra ejecución de Single-HYDFIT, pero minimizando Δ (ec. III.35) en función de p y a , y b de manera indirecta mediante la relación $b = a/p$. En la Figura III.16 se muestra el gráfico de contorno obtenido, Δ en función de p y a . De nuevo, obtenemos dos mínimos en el rango empleado para el barrido. Los resultados se muestran en la Tabla III.7 (entrada “Elipsoide”), que son las dimensiones de los elipsoides finalmente empleadas para construir nuestro modelo de “grano grueso”. Es necesario resaltar que los resultados se asemejan bastante a los obtenidos en su día por Carrasco *et al* [76, 77], pero han sido obtenidos mediante una vía más robusta y sistemática y los consideramos más precisos.

Respecto a las propiedades experimentales, no las hemos empleado *a priori* en este caso para hallar los elipsoides hidrodinámicos equivalentes (por haber un conjunto de datos insuficiente, como indicábamos anteriormente), pero podemos realizar una comparación *a posteriori* y deducir si son compatibles con los resultados obtenidos. En la Tabla III.7 se muestran los valores experimentales de R_g y s respecto a los del elipsoide adoptado como

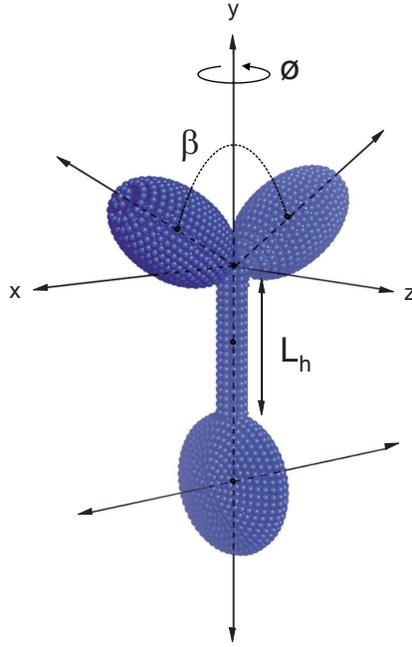


Figura III.17: Representación del modelo de esferas empleado por HYDROSUB, junto a los parámetros geométricos que definen el modelo.

modelo. La diferencia típica está alrededor del 5%, coincidiendo con el error experimental considerado para calcular las propiedades.

El modelo genérico del anticuerpo IgG3 consiste, por tanto en: (a), los dos elipsoides prolatos que representan los dominios Fab, con semiejes de $a_{ab} = 43.7 \text{ \AA}$ y $b_{ab} = 24.9 \text{ \AA}$; (b), el elipsoide oblato para el dominio Fc, con semiejes de $a_c = 14.5 \text{ \AA}$ y $b_c = 43.0 \text{ \AA}$; y (c), la región bisagra, representada por una varilla de longitud L_h - es esta, junto a alguna otra variable estructural, para distintas formas de IgG3, lo que nos queda por determinar. Las propiedades para esta composición de subunidades son calculadas utilizando HYDROSUB [25], donde la información estructural es un conjunto de datos; para cada subunidad las coordenadas de sus centros, $(x_i, y_i, z_i)_b$, y los ángulos polares y azimutales, θ_i y ϕ_i , que determinan la orientación de los ejes de revolución principales, todo referido a un sistema de referencia convenientemente elegido. Como se muestra en la Figura III.17, el cilindro equivalente a la región bisagra está situado en la parte negativa del eje z , cuyo centro está situado en las coordenadas $(x_h, y_h, z_h) = (0, 0, L_h/2)$, y su orientación, $\theta_h = \phi_h = 0$. Al final del cilindro está situada la subunidad Fc, centrado a $(x_c, y_c, z_c) = (0, 0, L/2 + a_c)$, y su plano ecuatorial en el plano (y, z) , con $\theta_c = 90^\circ$ y $\phi_c = 0^\circ$. Los dos elipsoides prolatos, equivalentes a las subunidades Fab, están situados en el otro extremo del cilindro. A la hora

de diseñar el barrido conformacional, para reducir el número de parámetros, decidimos que el ángulo entre ambas subunidades variara manteniendo su simetría, definiendo un solo ángulo, β , biseccionado en el eje z , de tal manera que los ángulos polares están en la forma $\phi_{ab1} = \phi$ y $\phi_{ab2} = \phi + 180^\circ$. Entonces, las coordenadas de sus centros serán $(x_{ab}, y_{ab}, z_{ab}) = (\pm a_{ab} \sin \theta_{ab} \cos \phi, \pm a_{ab} \sin \theta_{ab} \sin \phi, \pm a_{ab} \cos \theta_{ab})$. Nótese que ϕ es el ángulo entre el plano ecuatorial de Fc y el plano que definen los ejes principales de revolución de las subunidades Fab.

En resumen, nuestro modelo de la estructura global de las variantes del anticuerpo IgG contiene tres parámetros a determinar. La longitud de la región bisagra, L_h , el ángulo entre las subunidades Fab, β , y el ángulo que define la orientación de las mismas, ϕ . Nuestra intención es determinar la estructura global de tres moléculas de IgG3 distintas, el anticuerpo nativo IgG3wt, y las variantes IgG3m15 e IgG3HM5. En la amplia bibliografía disponible para estos y otros anticuerpos, existen datos publicados de sus propiedades en disolución: coeficiente de sedimentación, s , viscosidad intrínseca, $[\eta]$, radio de giro, obtenido mediante dispersión de rayos X, R_g y distancia máxima, D_{max} . Los datos experimentales disponibles están resumidos en la Tabla III.8.

De acuerdo a nuestra estrategia de cálculo multiescalar y barrido conformacional (Figura III.17), la estructura que mejor ajusta es obtenida como ya hemos indicado, con un análisis final mediante HYDROFIT [75] de los datos de las propiedades en disolución calculadas por HYDROSUB [25], a partir de un número elevado de estructuras (6156) generadas por MULTISUB [33]. Mediante una programación sencilla, se adaptó nuestro problema al código genérico de MULTISUB, el cual, teniendo en cuenta las dimensiones, orientaciones y los centros de las subunidades, genera las estructuras, que son escritas en archivos de salida. Estos a su vez, son utilizados como entrada para HYDROSUB, el cual realiza los cálculos de sus propiedades en disolución.

Antes de proceder a la etapa final (optimización) de HYDROFIT, es conveniente realizar una evaluación preliminar de la dependencia de las propiedades en disolución (se han evaluado cuatro de ellas) predichas con los parámetros del modelo, L_h , β y ϕ . (Figura III.7). De los tres parámetros a evaluar, dos se mantienen constantes en un valor razonable, variando el tercer parámetro, representándose los valores de las propiedades en disolución relativas al valor máximo en el rango de la propiedad que varía. Encontramos que, mientras las propiedades varían apreciablemente con L_h y β , por otra parte son prácticamente insensibles a la variación del parámetro ϕ , ya que se mantienen constantes los centros y las dimensiones de las subunidades, variando tan solo la orientación de las mismas. Desde

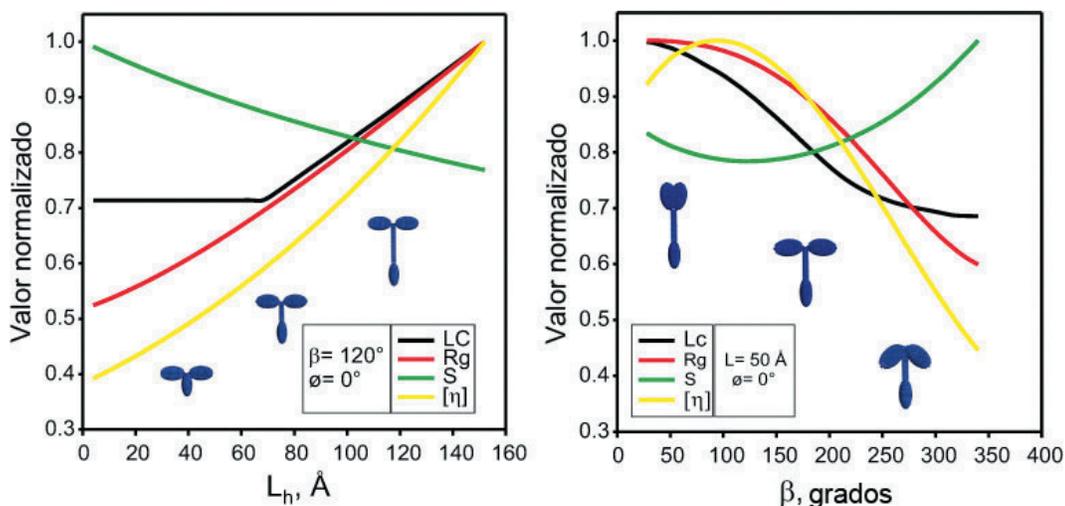


Figura III.7: Dependencia de las propiedades calculadas con los parámetros del modelo de IgG3. Dos de los tres parámetros se mantienen constantes, mientras que se varía el otro. Las propiedades son expresadas como relación entre el valor de la propiedad respecto al máximo obtenido en el rango donde el parámetro varía.

la teoría, si nos centramos en el R_g de las estructuras compuestas por subunidades [86], sabemos que el valor de esta propiedad depende de las distancias entre los centros y del radio de giro de cada subunidad, y para L_h y β fijos las distancias no varían, con ϕ , por lo cual el R_g no debe variar, y el resto de las propiedades se comportarán de manera similar con la variación de ϕ . Por consiguiente, y para reducir el número de estructuras a calcular, solo vamos a tener en cuenta los parámetros que producen una variación significativa en alguna de las propiedades a evaluar, L_h y β . Nótese que esta variación en las propiedades evaluadas puede llegar a ser considerable, cerca del 60 %, dentro del rango de longitudes L_h y ángulos β empleado (ver Figura III.7).

A continuación, MULTISUB genera los modelos, empleando para ello un algoritmo que calcula las estructuras a partir de L_h y β . Se generaron un total de 6156 modelos hidrodinámicos. MONTESUB e HYDROSUB realizan el cálculo de las propiedades en disolución de cada modelo. Y por último, HYDROFIT se encarga de la optimización de estos dos parámetros estructurales empleando, por un lado, las propiedades en disolución calculadas para todos los modelos generados, y por otro lado, tres conjuntos de propiedades experimentales, pertenecientes a las tres variedades de anticuerpo (tabla III.8). Anali-

zando la dependencia de Δ (ec. III.35) en función de L_h y β , obtenemos los gráficos de contorno que se muestran en la Figura III.8.

Resultados

Para el anticuerpo nativo IgG3wt (Figura III.8a), se obtiene un único mínimo que corresponde a una única estructura hallada en el ajuste de propiedades experimentales. Esta estructura presenta una larga región bisagra, con un L_h de $\approx 92 \text{ \AA}$, que confirma los datos obtenidos en un trabajo anterior [82]. Respecto al ángulo β , se obtiene una típica forma de “T”, con un ángulo de $\approx 200^\circ$. Para esta estructura, se obtiene un $\Delta = 0.008$, indicando que la diferencia típica entre los radios equivalentes calculados respecto a los experimentales es de tan solo un 0.8 %. La Tabla III.8 muestra las propiedades en disolución calculadas para esta estructura, con la diferencia típica con los valores experimentales.

Respecto a la variante IgG3m15, con una región bisagra parcialmente eliminada, los resultados son los mostrados en la Figura III.8b. Si inspeccionamos los resultados obtenidos, vemos que se obtiene un mínimo absoluto, y a diferencia del anterior, otros 2 mínimos relativos muy próximos al absoluto. Por consiguiente, obtenemos 3 estructuras posibles, las cuales son “numéricamente” válidas, pero deben descartarse las incompatibles con las evidencias experimentales. Proponemos la estructura que presenta un L_h de $\approx 36 \text{ \AA}$, y un ángulo β de 252° , por presentar una región bisagra compatible con la longitud esperada de su secuencia de aminoácidos, y un ángulo β muy próximo al esperado, similar a la de la variante nativa.

Por último, la variante HM5. Como sucedía en el caso anterior, HYDROFIT nos muestra la existencia de tres mínimos. Nótese que el aspecto de la función de optimización para la variante HM5 (Figura III.8c) es muy similar a la obtenida para m15 (Figura III.8b), debido principalmente a que ambas variantes presentan unas propiedades en disolución casi idénticas. Además de no disponer de datos de viscosidad intrínseca para esta variante (ver Tabla III.8), observamos algunas diferencias, como que el mínimo del modelo hallado para m15, aparece en este caso menos definido. El mínimo absoluto, queda situado en la región $L_h = 0 \text{ \AA}$, $\beta = 40^\circ$, siendo compatible con los resultados esperados, al ser la única conformación que no presenta región bisagra. Además, podemos observar que el ángulo entre los fragmentos Fab es relativamente pequeño y el modelo adopta una forma de “Y”.

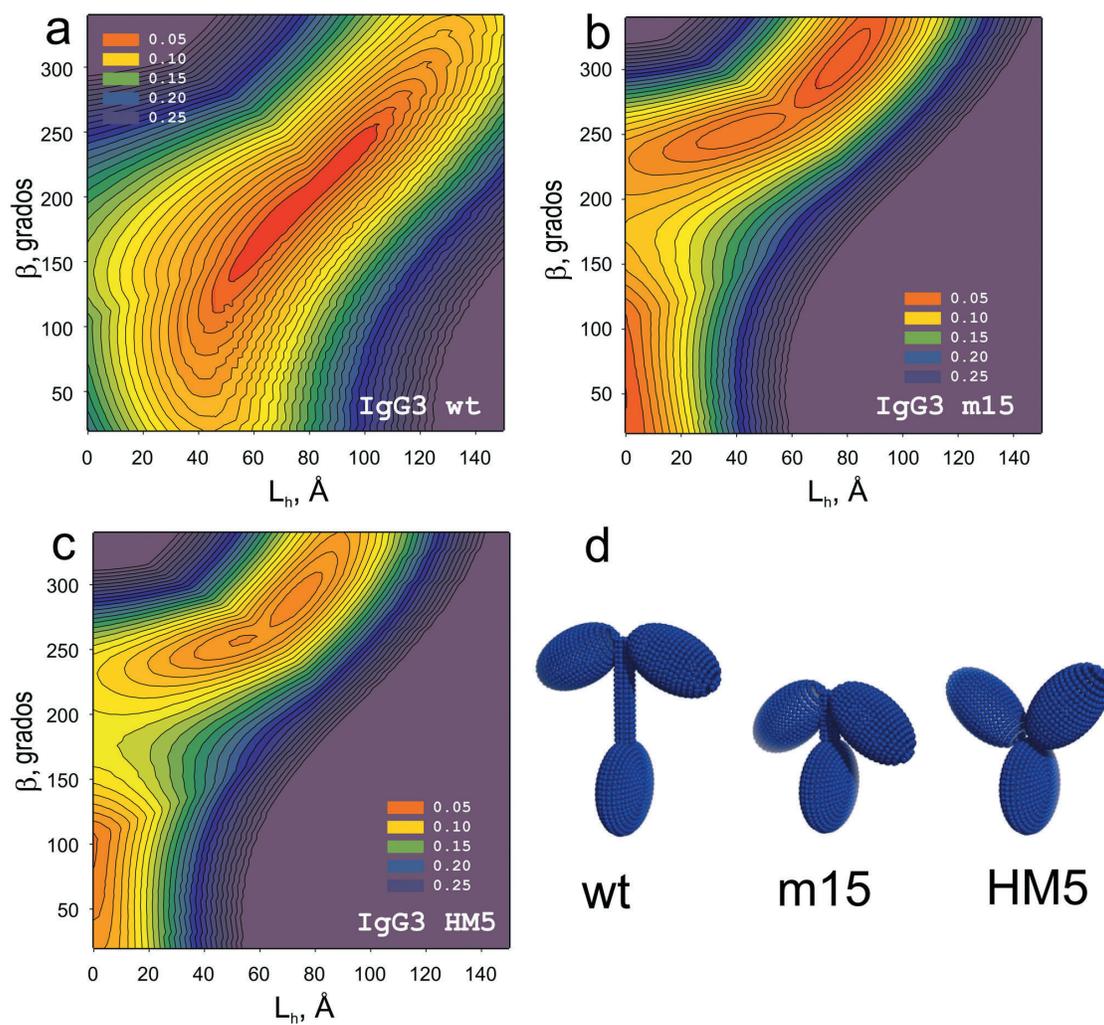


Figura III.8: Mapa de contorno de la función objetivo Δ frente a los parámetros del modelo de anticuerpo. Las figuras en (d) corresponden a los modelos que consideramos que mejor ajuste presentan en cada caso a) IgG3wt. (b) IgG3m15. (c) IgG3HM5. (d) Vista de los tres modelos obtenidos.

Propiedad	IgG3wt	IgG3m15	IgG3HM5
M , Da	156943	146771	143529
\bar{v} , cm ³ g ⁻¹	0.725	0.726	0.727
$S_{20,w}^0$, S (exptal)	6.11	6.77	6.68
Rg , Å (exptal)	71.6	54.2	54.8
$[\eta]$, cm ³ g ⁻¹ (exptal)	9.9	5.7	-
D_{max} , Å (exptal)	195	165	163
L_h , Å	92	36	0
β , grados	212°	252°	96°
$S_{20,w}^0$, S / % dif	6.11 / -0.03	6.59 / 2.71	6.27 / 6.10
Rg , Å / % dif	70.8 / 1.17	51.4 / 5.16	53.0 / 3.29
$[\eta]$, cm ³ g ⁻¹ / % dif	10.2 / -3.0	6.32 / -10.9	6.91 / -
D_{max} , Å / % dif	196 / -0.51	168 / -1.8	166 / -1.84
Δ Valor de la función (%)	0.791	4.338	4.806

Tabla III.8: Parámetros experimentales empleados para cada uno de los modelos de IgG utilizado en este estudio. Se expresan también los resultados obtenidos mediante nuestro modelo, y los errores con respecto a los experimentales.

Las tres estructuras se asemejan notablemente con las halladas para wt, m15 y HM5 por Lu et al [82] en una reciente publicación, aunque empleando este un procedimiento mucho más laborioso y menos sistemático. Además, nuestro procedimiento detecta, en los mínimos locales, la posibilidad de estructuras alternativas.

Conclusiones

Se ha mostrado en esta parte de la presente tesis la posibilidad de abordar, de manera modesta, el problema “inverso” de la determinación estructural, al menos en aspectos globales, con base en las otras metodologías de nuestros cálculos de propiedades en disolución. Además, se muestra la estrategia multi-escala, empleando desde información a nivel atómico hasta modelos de “grano grueso”. El ejemplo relativo a la diferenciación estructural del IgG3 nativo y los dos mutantes pone de manifiesto cómo esas diferencias son detectables mediante experimentación y cálculo de propiedades en disolución, gracias a los métodos de programación de ese “problema inverso”, ya implementado en *Single-HYDFIT*, y *Multi-HYDFIT*, y ahora en el nuevo programa *HYDROFIT*.

Capítulo IV

Macromoléculas vermiformes

IV.1. Introducción

Existe una gran variedad de polímeros sintéticos y macromoléculas que poseen un esqueleto semiflexible, y que pueden ser representados por el modelo de cadena vermiforme [87]. De acuerdo con la definición estándar, el modelo vermiforme se emplea para representar ‘una macromolécula lineal compuesta por una cadena infinitamente delgada de curvatura continua; la dirección de la curvatura en un punto determinado es aleatoria’. El modelo describe un amplio espectro de cadenas con diferentes grados de rigidez, desde varillas rígidas hasta cadenas completamente flexibles, aunque su utilidad se centra en cadenas semiflexibles. En la literatura, el modelo es referido en ocasiones como la cadena de Kratky-Porod [88].

El modelo vermiforme es especialmente importante en biopolímeros. La doble cadena en hélice del ADN es el ejemplo más paradigmático [89, 90], pero existen otras aplicaciones también muy útiles, como polisacáridos (ver [91, 92]). Hace tres décadas, Yamakawa y Fujii presentaron un tratamiento teórico de las propiedades en disolución del modelo vermiforme [93, 94, 95]. Este tratamiento incluye diversas aproximaciones, siendo la más destacada el reemplazo del tensor de Oseen por su valor pre-promediado. Se ha visto que esta aproximación induce errores apreciables en los dos extremos del modelo: en los límites de la varilla rígida [53] y la cadena completamente flexible [96]. Por otro lado, Zimm y otros [97, 14, 98], propusieron el empleo del método de Monte Carlo de cuerpo rígido (RBMC). Este método, inspirado en el trabajo pionero de Kirkwood y Riseman [41], modela la macromolécula como una cadena de esferas. En el tratamiento de cuerpo rígido,

las cadenas son generadas por Monte Carlo, y para cada conformación, las propiedades son calculadas individualmente, empleando procedimientos rigurosos de cálculo hidrodinámico [8], para ser promediadas al final del proceso. Se ha comprobado que este tratamiento predice correctamente los resultados experimentales para cadenas completamente flexibles en condiciones ideales y de buen disolvente [96, 99, 100].

El tratamiento RBMC fue empleado por primera vez en cadenas vermiformes por Hagerman y Zimm [14] en su estudio de la dinámica rotacional de fragmentos de tamaño corto y medio de ADN, cuyas conformaciones están muy próximas a la varilla rígida. Este trabajo se hizo extensivo a otras propiedades en disolución (difusión traslacional y viscosidad intrínseca) por Molina et al [101]. Debido al alto coste computacional del método, su aplicabilidad quedó limitada por la potencia de cálculo disponible en aquellos años. En este capítulo, presentamos un nuevo trabajo basado en el método de Monte Carlo, donde estudiaremos las propiedades en disolución de los polímeros vermiformes, empleando los recientes avances en hidrodinámica macromolecular para el cálculo de propiedades [23], y el tratamiento de los resultados y los datos experimentales mediante los radios equivalentes y las relaciones de radios (sección II.1.3 [21]). El cálculo cubre el espectro completo del modelo vermiforme, desde partículas cilíndricas rígidas hasta cadenas completamente flexibles, cortas y largas. La relevancia se ilustra en el análisis de los datos experimentales de ADN - el ejemplo de molécula vermiforme más paradigmático - cubriendo un rango extremadamente amplio de tamaños moleculares, desde 10 a 10^5 pares de bases. Nuestras metodologías se ven facilitadas por el análisis de datos experimentales empleando la herramienta `Multi-HYDFIT` [21], donde han sido incorporados nuestros resultados de la simulación. Hemos observado que esta fácil y simple herramienta permite un ajuste rápido de las propiedades del ADN, en el amplio rango de pesos moleculares, obteniendo los parámetros del modelo vermiforme para la doble hélice de una manera muy precisa.

IV.2. Metodología

IV.2.1. Modelo vermiforme

La cadena vermiforme es un modelo donde la flexibilidad es continua a lo largo de la cadena polimérica, y no está centrada en puntos concretos. Este modelo ha sido ampliamente utilizado para la descripción de la física de polímeros semiflexibles, tanto sintéticos como de origen biológico. Dos características básicas del modelo vermiforme son la lon-

gitud de contorno y la longitud de persistencia. Ambos parámetros son importantes para describir la flexibilidad del modelo:

- La longitud de persistencia a se define, según la IUPAC, como la proyección media del vector extremo-extremo sobre el eje tangente del contorno de la cadena en el origen, en el límite de longitud infinita de la cadena. También puede ser expresado como la distancia sobre la cual la cadena persiste en la dirección de un determinado enlace b_j . Algunos autores prefieren emplear la *longitud del segmento estadístico de Kuhn*, λ^{-1} , cuya equivalencia con a es $a = \lambda^{-1}/2$.
- La longitud de contorno L corresponde a la longitud total extremo-extremo de la cadena si todos los elementos de la cadena estuvieran dispuestos de forma recta o estirada.

Este modelo es una aproximación a la estructura de la macromolécula, y no tiene en cuenta la estructura química de la misma. Por lo tanto, cuando hablamos de enlace, no nos referimos a un enlace químico, sino a una unidad rígida que une dos elementos N consecutivos. El modelo vermiforme presenta una teoría muy desarrollada y se conocen, gracias a diversos investigadores, las ecuaciones que determinan los valores promedios de sus magnitudes configuracionales, tales como $\langle s^2 \rangle$ y $\langle r^2 \rangle$.

El valor cuadrático medio de la distancia extremo-extremo, fue obtenido por Kratky-Porod en 1949 [87]:

$$\langle r^2 \rangle = 2aL \left[\left(1 - \frac{a}{L}\right) (1 - e^{-L/a}) \right] \quad (\text{IV.1})$$

Por otro lado, la ecuación relativa al valor cuadrático medio del radio de giro, atribuida al trabajo de Benoit-Doty en 1953 [102]:

$$\langle s^2 \rangle = 2aL \left[\frac{1}{6} - \frac{a}{2L} + \frac{a^2}{L^2} - \frac{a^3}{L^3} (1 - e^{-L/a}) \right] \quad (\text{IV.2})$$

Estas ecuaciones nos resultarán muy útiles para la determinación de las propiedades en disolución, como se tratará más adelante.

IV.2.2. Monte Carlo de cuerpo rígido (RBMC) y cálculo de propiedades de modelos de esferas

Tal y como se ha descrito en la introducción de este capítulo, las propiedades han sido calculadas desde la simulación Monte Carlo, en términos del tratamiento del cuerpo

rígido de entidades flexibles, donde las propiedades son obtenidas a partir de promedios conformacionales sobre las propiedades evaluadas para cada conformación individual, consideradas como partículas instantáneamente rígidas. El procedimiento requiere el cálculo hidrodinámico de cuerpo rígido, realizado sobre modelos de esferas que representan la cadena vermiforme, cuya construcción se detalla posteriormente. Para dicho cálculo, se ha empleado una metodología avanzada aplicada a la hidrodinámica de modelos de esferas [10, 68, 103], implementada en la herramienta computacional HYDRO++, que incluye como novedad diversas mejoras en el cálculo de coeficientes rotacionales y viscosidad intrínseca, haciendo hincapié en esta última, que incluye la denominada corrección de volumen ajustada (ver sección III.2).

Con el propósito de analizar convenientemente los resultados de la simulación, con el objetivo de encontrar la dependencia con los parámetros estructurales y obtenerlos a partir de datos experimentales, consideramos de gran utilidad el empleo del concepto de radios equivalentes y relaciones de radios [21], que han sido expuestos en la sección II.1.3). Reproducimos aquí los empleados para el presente problema: Los radios equivalentes para el radio de giro, para los coeficientes traslacionales y la viscosidad intrínseca:

$$a_G = \sqrt{\frac{3}{5}} R_g \quad (\text{IV.3})$$

$$a_T = \frac{k_B T}{6\pi\eta_0 D_t} \quad (\text{IV.4})$$

$$a_I = \left(\frac{3M[\eta]}{10\pi N_A} \right)^{1/3} \quad (\text{IV.5})$$

En las ecuaciones anteriores, η_0 y ρ son la viscosidad y la densidad del disolvente respectivamente, \bar{v} es el volumen específico parcial de la macromolécula, M es el peso molecular, T es la temperatura absoluta, y k_B y N_A son las constantes de Boltzmann y Avogadro respectivamente. Los radios a_G , a_T y a_I no dependen de estas magnitudes; dependen sólo del tamaño y conformación del soluto macromolecular, que en el caso del modelo vermiforme se determinan por sus parámetros estructurales L , a y d . Y estos radios pueden, a su vez, combinarse en las denominadas relaciones de radios:

$$GT = \frac{a_G}{a_T} = \frac{6\pi\eta_0}{f_t} \frac{R_g}{(3/5)^{1/2}} = \frac{6\pi\eta_0 D_t}{k_B T} \frac{R_g}{(3/5)^{1/2}} = \frac{6\pi\eta_0 N_A s}{M(1 - \bar{v}\rho)} \frac{R_g}{(3/5)^{1/2}} \quad (\text{IV.6})$$

$$GI = \frac{a_G}{a_I} = \left(\frac{10\pi N_A}{3[\eta]M} \right)^{1/3} \left(\frac{5}{3} \right)^{1/2} R_g \quad (\text{IV.7})$$

$$IT = \frac{a_I}{a_T} = \frac{GT}{GI} = \left(\frac{3[\eta]M}{10\pi N_A} \right)^{1/3} \left(\frac{k_B T}{6\pi\eta_0 D_t} \right) \quad (\text{IV.8})$$

Recordamos que las relaciones de radios no dependen del tamaño absoluto, y están relacionadas exclusivamente con la forma o conformación del soluto. De nuevo, en el caso de la cadena vermiforme, estas relaciones de radios GT , GI e IT , son funciones de las magnitudes que determinan su conformación, es decir, los cocientes de sus parámetros estructurales, como pueden ser L/d y L/a (o cualquier combinación entre ellos), pero no de los valores individuales de L , a ó d . En la simulación de Monte Carlo, debemos asignar un valor arbitrario a cualquiera de estos parámetros de longitud, (por ejemplo, $d=20 \text{ \AA}$), y los otros dos parámetros varían de una manera conveniente. Los resultados de R_g , D_t y $[\eta]$ son transformados a relaciones GT , GI e IT , cuyas dependencias con los cocientes L/d y L/a (o cualquier función de ellos) no dependerán de la elección de d , por lo que las relaciones de radios tienen una validez universal para cualquier cadena o polímero vermiforme.

IV.2.3. Versión discreta del modelo vermiforme. La cadena de Schellman

La simulación del modelo vermiforme discreto se lleva a cabo a partir del modelo ϕ -isotrópico de Schellman. La validez de esta aproximación quedó demostrada en los trabajos de Hagerman and Zimm [14] y García Molina [101] entre otros. Este modelo, en un principio, fue propuesto por Schellman [104] para describir la flexibilidad del ADN y supone una cadena formada por N esferas tangentes, de radio σ , que representan los puntos de fricción con el disolvente. Están unidas por $N - 1$ enlaces de longitud b . La longitud de contorno de la cadena (o longitud del cilindro equivalente) es $L = 2\sigma N$ o bien, $L = Nb$. En la Figura IV.1 podemos observar una representación del modelo.

La configuración de una cadena de Schellman viene dada por un conjunto de ángulos θ_i y ϕ_i (entre segmentos consecutivos), cuyos valores dependerán de los parámetros del modelo, tales como longitud de segmento b , longitud de persistencia a y longitud de contorno L . La rigidez queda reflejada en los valores de θ_i , asumiéndose que los ángulos ϕ_i

son isotrópicos; sus valores responden a una distribución uniforme en $(0, 2\pi)$, calculándose a partir de:

$$\phi_i = 2\pi RAN \quad (IV.9)$$

donde RAN es un valor aleatorio de distribución uniforme en el intervalo $(0, 1)$, proporcionado por un generador de números pseudoaleatorios.

Respecto al ángulo θ_i , según el modelo de Schellman se obtiene mediante una aplicación sucesiva de dos desplazamientos angular ortogonal ($\theta_{i,1}$ y $\theta_{i,2}$) en los planos XY y XZ de la configuración $i - 1$. La expresión para θ_i vendrá dada por:

$$\theta_i \cong \sqrt{(\theta_{i,1}^2 + \theta_{i,2}^2)} \quad , \text{ para } \theta_i < 1 \text{ rad} \quad (IV.10)$$

En este caso, los ángulos $\theta_{i,j}$ siguen una distribución normal o gaussiana, centrados en $\theta_{i,j} = 0$, con una desviación estándar de $\sqrt{(b/a)}$. Los valores son calculados a partir de un generador de números aleatorios de distribución normal, mientras que las coordenadas cartesianas de las esferas se calculan a partir de estos ángulos empleando métodos estándar [105]. No debemos obviar que esta versión discreta se trata de una aproximación, y que únicamente obtendremos una verdadera cadena vermiforme en el límite de N infinito.

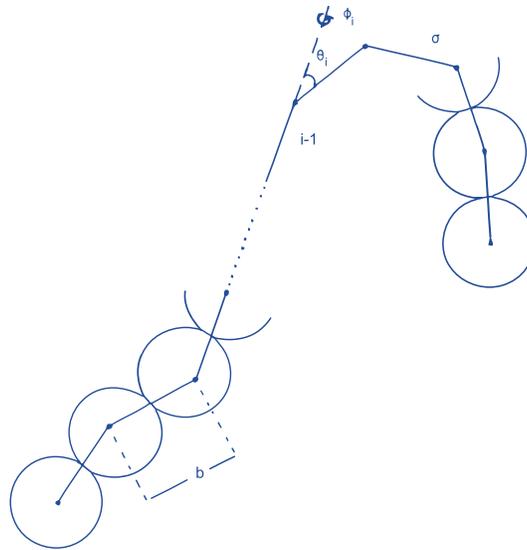


Figura IV.1: Esquema del modelo de Schellman [104]

Para validar el modelo, debemos comprobar que los promedios obtenidos en la simulación se corresponden a los esperados, en el caso de alguna propiedad calculable de antemano, como es el radio de giro, dado por la ecuación IV.2.

Para todo lo que sigue, conviene definir un nuevo parámetro, $X = L/a$, que se obtiene de dividir la longitud total de la cadena L entre la longitud de persistencia a . De esta manera, si todas las cadenas generadas tienen la misma longitud, la rigidez de la cadena será máxima (será una varilla) cuando X se aproxima a 0, y será prácticamente una cadena gaussiana, en el límite flexible, cuando $X \rightarrow \infty$. De esta manera, podemos generar cadenas de N elementos y realizar un barrido y hallar los promedios para cada X y N . Se realizaron 6 series de simulaciones para distintos N entre 5 y 500 elementos, promediando 2000 cadenas en cada caso. Los resultados aparecen en la figura IV.2.

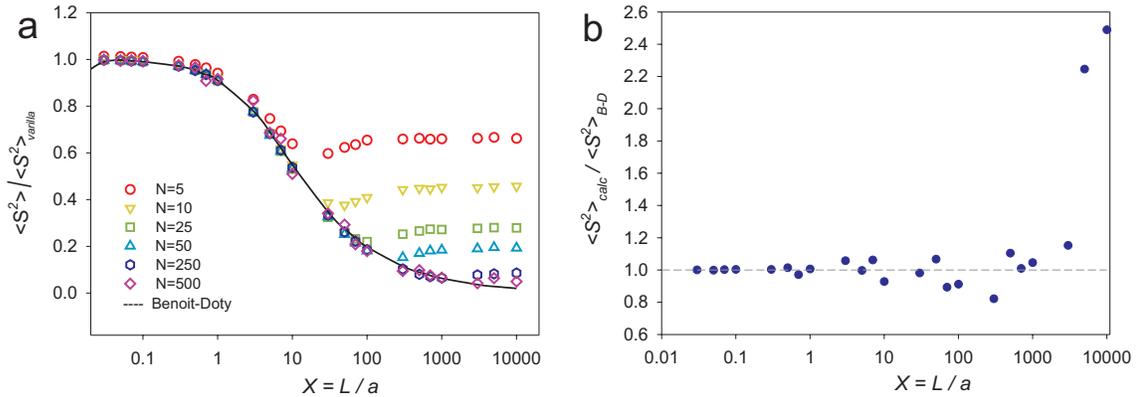


Figura IV.2: **(a)** Representación del promedio del radio de giro cuadrático, normalizado por conveniencia al radio de giro cuadrático de una varilla de igual longitud de contorno, frente al parámetro X , para cadenas formadas por N elementos, promediando un total de 2000 cadenas para cada caso ($n_{MC} = 2000$). **(b)** Representación de la relación del promedio del radio de giro cuadrático calculado ($N = 500$) respecto al hallado mediante la teoría de Benoit-Doty [102], de nuevo frente al parámetro X .

Vemos cómo la validez del modelo discreto se extiende hasta valores de X tanto mayores cuanto mayor es N . Apreciamos a simple vista, que para $N = 5$, 50 o 500, el límite de validez de X es, aproximadamente, 10, 100 o 1000 respectivamente. Esto nos sugiere que el valor máximo de X es $2N$. Como $X = L/a$ y $N = L/b$, definimos otra relación adimensional,

$$Z = \frac{X}{N} = \frac{b}{a} \quad (\text{IV.11})$$

llegamos a la conclusión de que el rango de validez del modelo discreto de Schellman es para aproximadamente $Z \leq 0.5$.

El modelo vermiforme incluye como casos límite el cilindro o varilla rígida y la cadena completamente flexible. Como asuntos previos a la descripción (modelo, teoría y simulación) del modelo vermiforme, vamos a repasar aspectos de esos casos límite.

IV.2.4. La cadena vermiforme en el límite flexible

La rigidez local de la cadena se mide por la longitud de persistencia a , pero la conformación global de la misma depende también de su longitud de contorno L . En el límite de cadenas muy largas, $L/a \rightarrow \infty$, y $N \rightarrow \infty$, la cadena vermiforme se reduce a la cadena completamente flexible, (*random coil*) donde las propiedades conformacionales e hidrodinámicas son bien conocidas. El modelo de cadena completamente flexible puede adoptar diversas concreciones, dependiendo de cómo se represente la estructura local, de cómo se exprese su tamaño o longitud, etc. Pero globalmente, el modelo posee unas leyes generales. Así, la distancia extremo-extremo cuadrática media, $\langle r^2 \rangle$, es proporcional a ese tamaño, expresado como número de elementos, N , longitud de contorno, L , o peso molecular, M [106]:

$$\langle r^2 \rangle \propto N \propto L \propto M \quad (\text{IV.12})$$

y el radio de giro cuadrático medio viene dado por

$$\langle s^2 \rangle = \frac{\langle r^2 \rangle}{6} \quad (\text{IV.13})$$

de manera que el radio de giro para cualquier modelo que tenga cierta flexibilidad, que se obtiene como

$$R_g = \sqrt{\langle s^2 \rangle} \quad (\text{IV.14})$$

tiene una dependencia con el tamaño de la cadena de la forma:

$$R_g \propto N^{1/2} \propto L^{1/2} \propto M^{1/2} \quad (\text{IV.15})$$

Estas relaciones son válidas para una cadena completamente al azar en la que (aparte de interacciones de corto alcance), la disposición de dos segmentos suficientemente separados, de uno respecto a otro, es totalmente aleatoria. Lo anterior solo sería válido si hubiera un balance entre repulsiones y atracciones que solo ocurre en unas condiciones de disolvente y temperatura especiales, denominadas condiciones *theta*. También se cumple esto aproximadamente cuando las cadenas son relativamente cortas, de manera que esos contactos de largo alcance son infrecuentes.

La relación entre radio de giro cuadrático medio y distancia extremo-extremo cuadrática media, ecuación IV.13 se cumple, no de manera exacta, pero sí como una muy buena

aproximación incluso fuera de las condiciones *theta*. Los aspectos básicos de la estadística conformacional de la cadena flexible pueden encontrarse en numerosos libros de texto [43, 105].

Las propiedades hidrodinámicas, como el coeficiente de fricción traslacional, f_t , (determinable a partir de los coeficientes de sedimentación s y difusión traslacional D_t), y la viscosidad intrínseca $[\eta]$, se han relacionado tradicionalmente con el radio de giro mediante las denominadas constantes de Flory [106, 107]:

$$P_0 = \frac{f}{\eta_0 \langle r^2 \rangle^{1/2}} \quad (\text{IV.16})$$

y

$$\Phi = \frac{[\eta]M}{\langle r^2 \rangle^{3/2}} \quad (\text{IV.17})$$

Reemplazando $\langle r^2 \rangle$ por $6R_g^2$, notamos como P_0 y Φ contienen las relaciones f_t/R_g y $[\eta]/R_g^3$, que a su vez aparecen en las relaciones GT y GI , en las ecuaciones IV.6 y IV.7. Dando sus valores a K_B y N_A , estas ecuaciones conducen a las equivalencias (para cualquier estructura): $P_0 = 9.93/GT$ y $\Phi = 9.23 \times 10^{23}/GI^3$. La determinación precisa de la constante *universal* de Flory para cadenas flexibles ha sido objeto de numerosos trabajos. Cálculos de Monte Carlo de cuerpo rígido (RBMC), propuestos por Zimm [97], y realizados por García de la Torre y colaboradores [96], dan como resultados $P_0 = 6.0$ y $\Phi = 2.53 \times 10^{23}$, en excelente concordancia con los datos experimentales del sistema poliestireno/ciclohexano/35°, ampliamente caracterizado en numerosos trabajos. Por consiguiente, los valores teóricos correspondientes de nuestras relaciones de radios, en el límite flexible y en ausencia de volumen excluido, son $GT=1.65$ y $GI=1.53$. Estos valores corresponden a los valores de $N \rightarrow \infty$ para cualquier Z , por lo que conociendo el valor límite, podemos estimar el valor de nuestras relaciones de radios en cualquier valor de N . Nótese que para facilitar la extrapolación desde los resultados de los modelos discretos de N finito (con un máximo de $N = 5000$), se toma por conveniencia la variable $1/\sqrt{N}$. En la Figura IV.5a, donde se muestra la variación del cociente GT , podemos observar cómo los valores numéricos tienden y extrapolan al valor teórico. Esta situación es similar para el cociente GI e IT , donde las representaciones gráficas tienen el mismo aspecto (Figura IV.5b y IV.5c). Del mismo modo, los resultados numéricos aparecen reflejados en la Tabla IV.1.

IV.2.5. Efecto del volumen excluido

El efecto del volumen excluido (VE) tiene una influencia importante en las propiedades en disolución de cadenas largas y flexibles, pues hay cierta repulsión neta entre segmentos, de manera que un segmento puede estar próximo pero no superponerse a otro. La influencia combinada de longitud y flexibilidad se expresa por el número de segmentos estadísticos (de Kuhn), $L/(2P) = X/2$. Para ciertos polímeros, donde el modelo vermiciforme es aplicable, si la longitud de persistencia a es elevada, y/o la longitud de contorno es relativamente corta, el número de segmentos estadísticos no será muy elevado, por lo que el efecto de volumen excluido puede ser ignorado. Sin embargo, se ha intentado que nuestro tratamiento cubra completamente los casos en el límite flexible del modelo vermiciforme, incluyendo cadenas muy largas. Hemos introducido el efecto del volumen excluido de una manera sencilla para cubrir estos casos poco frecuentes. La simulación Monte Carlo de la cadena SHZ fue realizada de nuevo, pero incluyendo la restricción de rechazar las cadenas que presenten solapamientos entre las esferas que las componen. Esta restricción limita los cálculos, de tal manera que no podemos cubrir todo el rango de Z a partir de un N elevado $N > 500$, debido a que la probabilidad de encontrar una cadena no-solapante en la zona flexible disminuye exponencialmente con N . Las relaciones de radios GT y GI calculadas con VE son ligeramente superiores a las calculadas en ausencia de VE, reflejándose el efecto expansor del volumen excluido. Sin embargo, el aspecto de la variación de GT y GI con N y Z es muy similar. Los resultados están tabulados en la Tabla IV.2 y en la Figuras IV.5d, IV.5e y IV.5f.

En el límite flexible, los valores consenso de los parámetros de Flory $P_0 = 5.3$ y $\Phi = 1.9 \times 10^{23}$ con VE, obtenidos por cálculos de Monte Carlo de cuerpo rígido [100] y en buena concordancia con los datos experimentales del sistema poliestireno/etilbenzeno/25°C [108, 109] y poliestireno/tolueno/20°C, se transforman como antes se indica en relaciones de radios hidrodinámicos [21], resultando los valores $GT = 1.87$ y $GI = 1.69$. Siguiendo la misma estrategia, estos valores son empleados para el límite $N \rightarrow \infty$ ó $N^{-1/2} = 0$. En presencia de volumen excluido, el radio de giro es usualmente caracterizado mediante el factor de expansión α_g :

$$\alpha_g^2 = R_{g,VE}^2 / R_g^2 = \langle s^2 \rangle_{VE} / \langle s^2 \rangle \quad (\text{IV.18})$$

Nótese que hemos empleado el subíndice VE para los valores en presencia de volumen excluido, y se omite el subíndice para los casos en los que este no interviene.

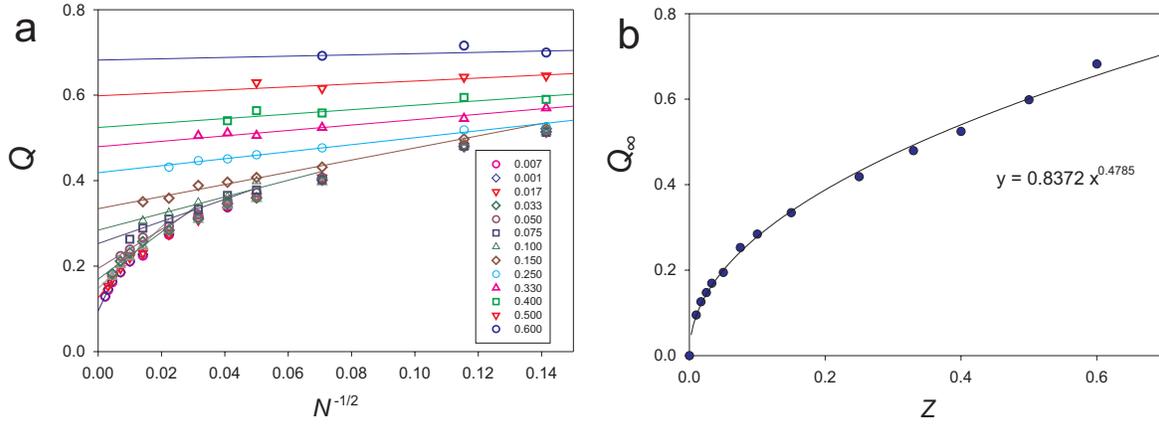


Figura IV.3: (a) Valores de $Q = \alpha_g^2/N^{0.17}$ vs. $N^{-1/2}$ para los valores indicados de Z . Las líneas rectas representan la extrapolación a longitud de cadena infinita. (b) Dependencia de los valores de Q_∞ con Z . La línea continua es el ajuste por mínimos cuadrados a una ley de potencia, $Q_\infty = 0.84 \times Z^{0.48}$.

La teoría para el efecto del volumen excluido es particularmente compleja en el caso de cadenas semiflexibles. En la mayoría de aplicaciones, el efecto VE no es apreciable; sin embargo este efecto es muy intenso para cadenas extremadamente largas y flexibles. Por ello, hemos adoptado una descripción simple para estos casos. Para cadenas muy largas, $R_g \propto Z^{1/2}$ y $R_{g,VE} \propto Z^\nu$, con $\nu \approx 0.6$, o siendo más precisos, $\nu = 0.585$ [105]. Por consiguiente, la expresión del factor de expansión será

$$\alpha_g^2 = QN^{0.17} \quad (\text{IV.19})$$

El valor del término Q depende de la fuerza del efecto VE, que en nuestro modelo está relacionada con el diámetro de la esfera, por lo que se trata del parámetro VE de esfera dura. Empleando nuestros resultados del cálculo de Monte Carlo para R_g and $R_{g,VE}$, evaluamos α_g^2 y los valores numéricos de $Q = \alpha_g^2/N^{0.17}$ (Tabla IV.2 y Figura IV.3a). En nuestras cadenas finitas, este valor presenta una ligera dependencia con N , y pueden ser extrapolados para N infinito, cuyo valor de Q en ese límite denominaremos en lo sucesivo Q_∞ . Si analizamos este valor de Q_∞ respecto al parámetro Z , cabe esperar que los valores de Q_∞ disminuyan con $Z = b/a$. Realizando un ajuste de mínimos cuadrados, obtenemos la ecuación empírica: (ley de potencia) $Q_\infty \propto 0.84N^{0.48}$.

IV.3. Simulación Monte Carlo

Como se describe en el apartado anterior, la configuración de una cadena de Schellman viene dada por un conjunto de ángulos θ_i y ϕ_i , cuyos valores dependerán de los parámetros del modelo, tales como longitud de segmento b , longitud de persistencia a y la longitud de contorno L . A continuación procederemos a la simulación de Monte Carlo de cadenas vermiformes y al cálculo de las propiedades conformacionales e hidrodinámicas del modelo. En primer lugar debemos establecer la configuración espacial de cada cadena generada. Para ello vamos a describir el algoritmo empleado en la generación de las cadenas.

IV.3.1. Generación de cadena

Con estas consideraciones pasamos a describir el algoritmo generador de cadenas vermiformes, con un determinado número de enlaces N con longitud b y una determinada longitud de persistencia a .

El primer paso de nuestro algoritmo consiste en situar espacialmente los tres primeros elementos de la cadena. El primer elemento se situará en el punto de origen del sistema de coordenadas:

$$x_1, y_1, z_1 = 0 \tag{IV.20}$$

El segundo elemento,

$$x_2, y_2 = 0 \quad , \quad z_2 = b \tag{IV.21}$$

y el tercero:

$$x_3 = 0 \quad , \quad y_3 = b \sin \theta \quad , \quad z_3 = b \cos \theta + b \tag{IV.22}$$

Respecto al resto de elementos, las coordenadas espaciales se calculan a partir de las tres anteriores, y los ángulos θ_i y ϕ_i generados según las ecuaciones IV.9 y IV.10. En lo sucesivo, y a efectos del volumen que ocupan, consideramos que los elementos de la cadena tienen forma esférica y son tangentes, como se indica en la figura IV.1, con un radio $\sigma = b/2$.

En lo expuesto hasta aquí, el modelo ignora el efecto del volumen excluido, que imposibilita que dos elementos de la cadena, separados a lo largo de su contorno, puedan no obstante estar excesivamente cercanos en el espacio, debido a la flexibilidad parcial de la cadena. En las teorías básicas de la cadena vermiforme [43, 87, 104] este efecto se ignora, por la dificultad de introducirlo teóricamente y porque para macromoléculas no muy flexibles, las aproximaciones tan cercanas son bastante poco probables. No obstante, no hay

ningún problema esencial (salvo un aumento del tiempo de computación) en introducir dicho efecto. Una manera de hacerlo sería rechazando aquellas conformaciones en las que la distancia entre los centros de las esferas cualesquiera sean inferior a un cierto valor, por ejemplo, 2σ . Pero a medida que aumenta el tamaño de la cadena y/o disminuye su rigidez, disminuye la probabilidad de generar una cadena completa que cumpla con dicha condición, por lo que el tiempo de computación nos limita en ese aspecto.

Desde las coordenadas, las propiedades de las conformaciones individuales pueden ser evaluadas. El radio de giro puede obtenerse fácilmente desde:

$$R_g^2 = \frac{1}{2N^2} \sum_{i=1}^N \sum_{j=1}^N \langle r_{ij}^2 \rangle \quad (\text{IV.23})$$

donde $\langle r_{ij}^2 \rangle$ es la distancia cuadrática promedio entre las bolas i y j .

IV.3.2. Procedimiento de simulación Monte Carlo

El modelo vermiforme discretizado se compone de un cierto número N de esferas tangentes y $N - 1$ enlaces, por lo tanto, todas tienen el mismo radio y la distancia entre los centros de las esferas (distancia de enlace b), es igual al diámetro de estas, es decir $b = 2\sigma$, siendo σ el radio de cada esfera.

El primer parámetro, de longitud adimensional, es el número de elementos, que ya hemos definido como N , que recordamos que se obtiene a partir de $N = L/2\sigma$, siendo L la longitud de contorno de la cadena. Este parámetro, equivale a la relación *longitud/grosor* de una cadena polimérica, y está íntimamente relacionado con el peso molecular M de la misma. No obstante, a efectos prácticos, consideramos conveniente emplear $N^{-1/2}$ en lugar de N para facilitar el tratamiento matemático a la hora de realizar las extrapolaciones, ya que, mientras $N \rightarrow \infty$, $N^{-1/2} \rightarrow 0$, y con una variación más suave que con N^{-1} . Nuestro objetivo incluye la simulación de cadenas poliméricas lineales de alto peso molecular, por lo que necesitamos alcanzar valores de N muy altos. Pero debemos tener en cuenta que, en el cálculo hidrodinámico de cada cadena generada, el tiempo de computación que se requiere, independientemente del tipo de máquina, va en función de N^3 , por lo que a medida que vamos generando cadenas cada vez más largas, y pesos moleculares mayores, el tiempo de computación se hace impracticable.

El segundo parámetro, Z , ya definido anteriormente (ec. IV.11), describe la flexibilidad, desde el límite de la varilla rígida $Z = 0$, hasta el límite de validez del modelo, que según

IV.3. SIMULACIÓN MONTE CARLO

hemos visto en el apartado anterior, lo hemos situado en $Z \approx 0.5$. Este parámetro equivale en este caso a la relación *grosor/longitud de persistencia*. A cada N le corresponde una serie de valores de Z . N describe la longitud de la cadena, desde los 3 elementos ($N = 3$) hasta los que permita el tiempo de computación, y Z el grado de flexibilidad. Estos dos únicos parámetros se han seleccionado para obtener una amplia gama de longitudes y flexibilidades en la simulación. En la Figura IV.4 podemos observar unos ejemplos de cadenas generadas, indicándose los parámetros N y Z que las caracterizan.

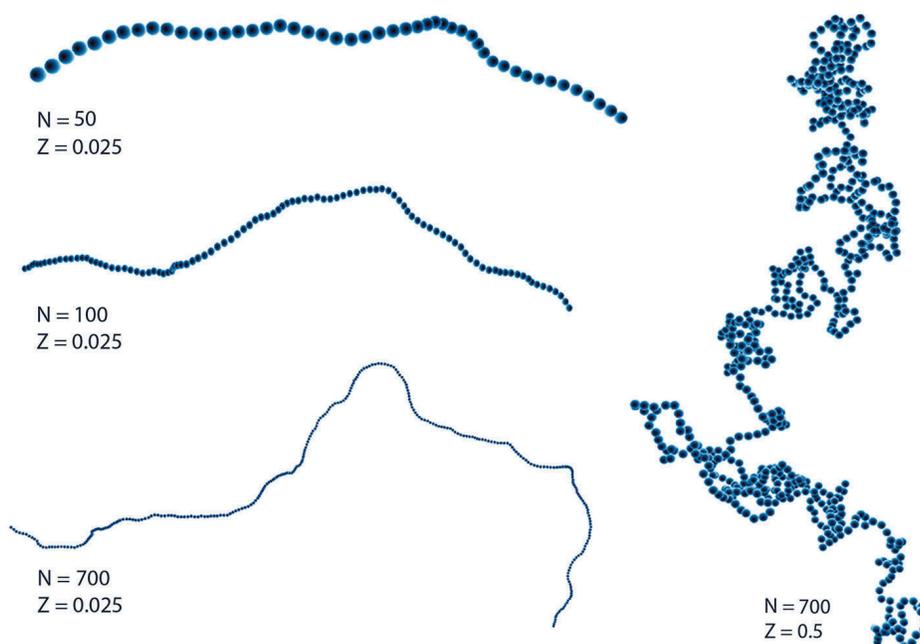


Figura IV.4: Ejemplos gráficos de cadenas generadas, indicándose los parámetros N y Z que las caracterizan.

El objetivo final de la simulación es obtener el promedio de las relaciones de radios XY para cada N y Z , cuyos valores dependerán únicamente de la conformación de las cadenas y de sus dimensiones relativas, y por consiguiente, no dependerán del tamaño ni de otras magnitudes absolutas, tales como el peso molecular (M) o el volumen total V , incluyendo las relativas al disolvente (temperatura, T , densidad, ρ o viscosidad, η_0). Los valores de las relaciones de radios XY , serán independientes de la elección que hagamos ‘*a priori*’ de estas magnitudes. Para obtener los resultados del cálculo hidrodinámico para cada N y Z , la muestra total se divide en 5 submuestras, y se calcula la media y la desviación

estándar de ellas para cada propiedad calculada. A continuación, se calculan los radios equivalentes para cada propiedad, según las ecuaciones IV.3, IV.4 y IV.5, y las relaciones de estos radios, según las ecuaciones IV.6, IV.7 y IV.8. Como resultado final obtenemos, para cada N y Z , los radios equivalentes y su error, aplicando las leyes de propagación de errores. Tras realizar las pruebas previas necesarias, tomamos un valor de compromiso para el tamaño muestral, teniendo en cuenta el error máximo (depende de N y Z) que se obtendrá en la simulación ($< 5\%$) y un tiempo de computación razonable. Los resultados finales se organizan en tablas de doble entrada, una para cada radio equivalente XY , con el efecto del volumen excluido (Tabla IV.1) y sin él (Tabla IV.2). El radio de giro calculado a partir de la ecuación de Benoit-Doty [102] (ec. IV.2) (en lo sucesivo, Rg_{BD}), podría diferir del promedio del calculado a partir de las coordenadas de las cadenas generadas en la simulación ($\langle Rg_{calc} \rangle$). Además del error propio de la simulación, se añaden otras desviaciones atribuibles al número finito de N . Por ello nos vemos obligados a añadir una corrección, a la que llamaremos RR , de tal manera que $RR = \langle Rg_{calc} \rangle / Rg_{BD}$ para cada valor de N y Z , y este valor será recogido en una tabla de doble entrada, junto a las tablas anteriores de XY (ver Tablas IV.1 y IV.2, y Figuras IV.5a, IV.5b y IV.5c). Estos datos, almacenados en ficheros de texto, los empleará Multi-HYDFIT para el cálculo de las propiedades en disolución.

IV.3.3. Predicción de las propiedades en disolución

Con los resultados de la simulación Monte Carlo, las propiedades en disolución se calculan a partir de sus parámetros estructurales L , a y M_L , junto con el peso molecular M de la muestra. En primer lugar, calculamos L a partir de la relación $M_L = M/L$ y hallamos el tamaño de esfera b a partir del diámetro d según $b = d/0.8165 = 1.225d$. Por consiguiente, $Z = b/a$ y $1/\sqrt{N} = \sqrt{b/L}$. A continuación, estos parámetros son empleados para hallar el Rg_{BD} (ec. IV.2). Por interpolación en las tablas de GT , GI , y RR obtenemos las relaciones correspondientes a cierto N y Z . Entonces, se obtienen los radios equivalentes $a_T = a_G/GT$ y $a_I = a_G/GI$, y finalmente D_t y $[\eta]$ empleando las ecuaciones IV.4 y IV.5.

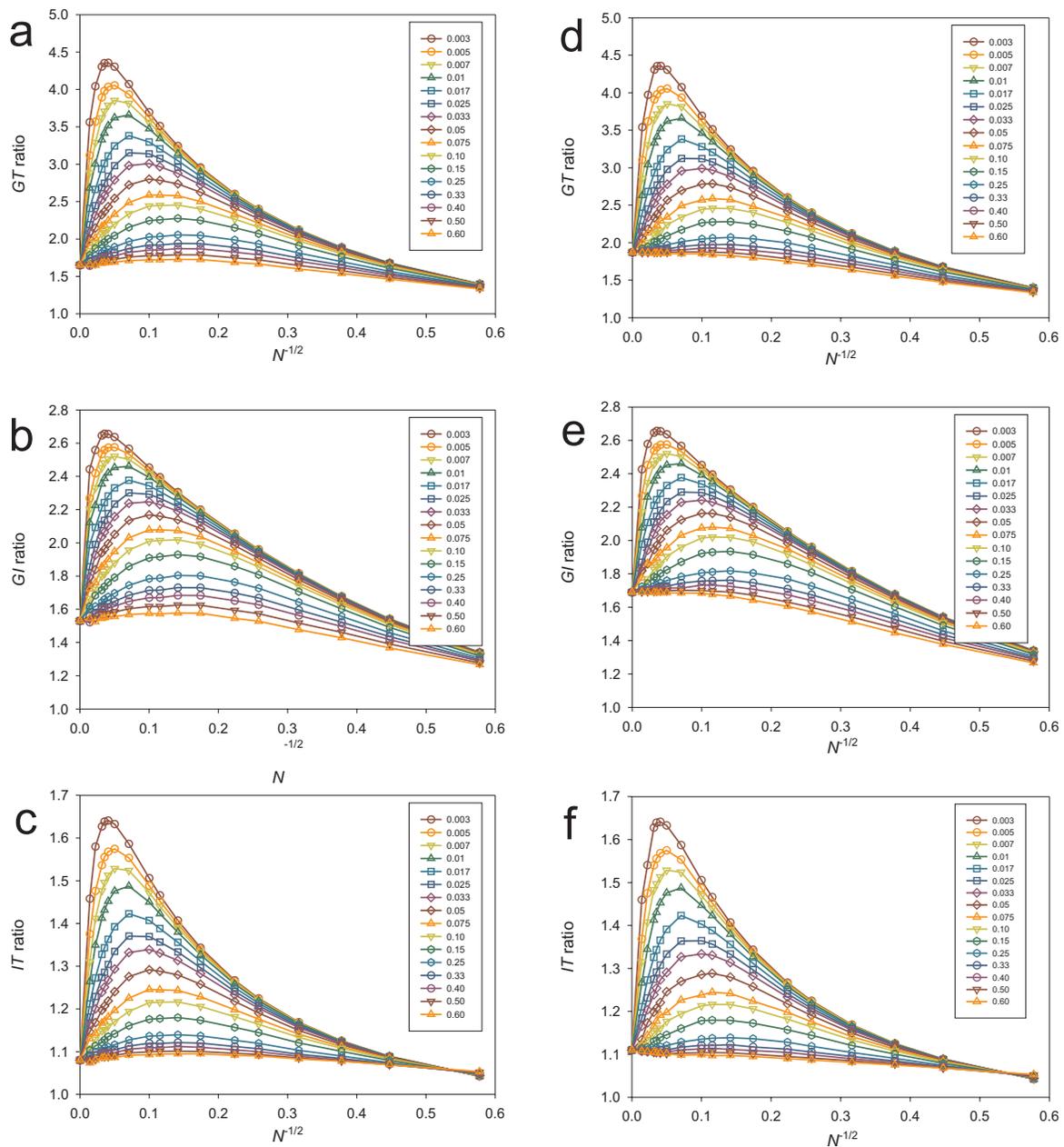


Figura IV.5: Representación de las relaciones de radios XY frente a los parámetros de la simulación N y Z . (a,b,c) Relación de radio GT , GI e IT frente a $N^{-1/2}$ en ausencia de volumen excluido. (d,e,f) Relación de radio GT , GI e IT frente a $N^{-1/2}$ en presencia de volumen excluido.

IV.4. Aplicaciones

IV.4.1. Determinación de los parámetros del modelo vermiforme desde sus datos experimentales

El objetivo último de nuestro estudio es implementar una metodología que permita extraer la información estructural relevante, contenida en los parámetros L , a y M_L , desde los datos experimentales. Este nuevo procedimiento para moléculas vermiformes ha sido implementado en el programa **Multi-HYDFIT**, ya publicado y descrito años atrás [21], complementando las teorías ya incluidas en el programa (teorías de Yamakawa-Fujii [93, 94, 95] y García Molina *et al* [101]).

Este programa, resulta de gran utilidad cuando, siendo un polímero lineal, disponemos de datos experimentales para una serie homóloga de muestras (o fracciones, según la técnica) de la misma molécula, pero con diferentes pesos moleculares. Para una serie de muestras de distinta masa M o longitud de contorno L , los parámetros estructurales a determinar serían el diámetro hidrodinámico d , la masa por unidad de longitud $M_L = M/L$, y la longitud de persistencia a . El programa **Multi-HYDFIT** considera estas tres constantes como parámetros ajustables y trata de optimizar la función Δ^2 multimuestra, empleando los conceptos de radios equivalentes y sus relaciones según la ecuación siguiente:

$$\Delta^2(a, M_L, d) = \frac{1}{N_s} \sum_{i=1}^{N_s} \left[\left(\sum_Y w_Y \right)^{-1} \sum_Y w_Y \left(\frac{a_Y(cal) - a_Y(exp)}{a_Y(exp)} \right)^2 \right] \quad (IV.24)$$

donde la sumatoria más externa se realiza sobre las N_s muestras, siendo Y cualquier propiedad observable, tales como Rg , D_t , s , $[\eta]$, etc. En la definición de Δ^2 empleamos desviaciones relativas de radios equivalentes a_Y , esto es, $(a_Y(cal) - a_Y(exp))/a_Y(exp)$. El empleo de las a_Y 's evita las diferencias en las sensibilidades de las diferentes propiedades con las dimensiones de la partícula, pero en el caso de análisis multimuestra el a_Y es mayor para las muestras más grandes, y esto se compensa utilizando la desviación relativa en lugar de las absolutas $a_Y(cal) - a_Y(exp)$, pues estas serían generalmente mayores para las muestras de mayor tamaño. Las optimizaciones pueden ser con pesos estadísticos o sin ellos.

En una primera etapa, **Multi-HYDFIT** emplea el procedimiento simple de rejilla, en el cual se indica un intervalo (menor/mayor) para cada parámetro, y se lleva a cabo una

búsqueda simple para el mínimo absoluto dando valores igualmente espaciados a cada parámetro dentro de su intervalo. Entonces, opcionalmente, en una segunda etapa se refina el mínimo mediante una minimización de Powell [110]. Además, este programa realiza un estimación de la incertidumbre de los valores óptimos de los parámetros, considerando los posibles errores experimentales. **Multi-HYDFIT** proporciona ficheros que pueden ser importados a los paquetes gráficos habituales.

La visualización (y la minimización en sí misma) puede ser difícil cuando están involucrados tres parámetros. De alguna manera, podemos fijar uno de los parámetros, siendo este de poca influencia en las propiedades (como puede ser el caso de d , en el caso de moléculas largas cilíndricas o vermiformes), o bien porque tenga un valor que está muy bien definido por su estructura macromolecular conocida (como puede ser para M_L).

Básicamente, y resumiendo el funcionamiento del programa: Supongamos que disponemos de datos experimentales para algunas de las propiedades siguientes: coeficiente de difusión traslacional, D_t ; coeficiente de sedimentación, s ; viscosidad intrínseca, $[\eta]$, uno de los tiempos de relajación, τ_x (generalmente el mayor, aunque también se emplea el tiempo de relajación armónico promedio), el radio de giro, R_g ; el covolumen, u ; y el volumen hidrodinámico de la partícula, V ; a partir de ellas, obtenemos un conjunto de radios equivalentes, a_X , y en su caso, las relaciones de radios. Los valores calculados, en este caso, se obtienen a partir de los datos extraídos de la simulación de Monte Carlo. La determinación estructural se llevará a cabo en un proceso de prueba y error, más concretamente mediante la optimización de la función Δ^2 (ec. IV.24) de desviación que establece la diferencia entre el conjunto de valores experimentales y los valores calculados para cada conformación.

CAPÍTULO IV. MACROMOLÉCULAS VERMIFORMES

Valores de GT (NO-EXCVOL)

N	$N^{-1/2}$	$Z= 0.003$	0.005	0.007	0.010	0.017	0.025	0.033	0.050	0.100	0.150	0.250	0.400	0.500
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.707	1.225	1.225	1.225	1.225	1.225	1.225	1.225	1.225	1.225	1.225	1.225	1.225	1.225
3	0.577	1.397	1.397	1.397	1.396	1.396	1.395	1.394	1.391	1.385	1.379	1.368	1.352	1.342
5	0.447	1.680	1.679	1.678	1.676	1.673	1.668	1.664	1.656	1.632	1.609	1.568	1.517	1.488
7	0.378	1.889	1.887	1.885	1.882	1.875	1.868	1.860	1.845	1.803	1.764	1.698	1.618	1.576
10	0.316	2.124	2.120	2.117	2.112	2.100	2.087	2.075	2.049	1.980	1.920	1.822	1.711	1.655
15	0.258	2.401	2.395	2.389	2.380	2.359	2.336	2.314	2.270	2.158	2.067	1.929	1.787	1.717
20	0.224	2.603	2.594	2.584	2.571	2.541	2.507	2.476	2.415	2.266	2.152	1.986	1.825	1.748
50	0.141	3.243	3.212	3.182	3.139	3.048	2.958	2.878	2.737	2.455	2.277	2.056	1.872	1.790
75	0.115	3.512	3.460	3.411	3.343	3.206	3.075	2.967	2.782	2.449	2.257	2.033	1.854	1.779
100	0.100	3.695	3.626	3.560	3.473	3.298	3.139	3.009	2.800	2.440	2.245	2.027	1.852	1.778
200	0.071	4.072	3.936	3.815	3.659	3.382	3.152	2.982	2.724	2.338	2.157	1.968	1.823	1.761
400	0.050	4.306	4.055	3.853	3.622	3.245	2.980	2.792	2.544	2.200	2.044	1.894	1.777	1.727
600	0.041	4.357	4.035	3.789	3.515	3.110	2.838	2.665	2.432	2.135	2.001	1.861	1.762	1.716
1000	0.032	4.304	3.893	3.620	3.324	2.935	2.686	2.524	2.322	2.066	1.925	1.812	1.754	1.710
2000	0.022	4.042	3.569	3.290	3.001	2.662	2.446	2.312	2.147	1.943	1.875	1.777	1.709	1.673
5000	0.014	3.562	3.120	2.886	2.680	2.410	2.191	2.089	2.001	1.846	1.783	1.745	1.641	1.662

Valores de GI (NO-EXCVOL)

N	$N^{-1/2}$	$Z= 0.003$	0.005	0.007	0.010	0.017	0.025	0.033	0.050	0.100	0.150	0.250	0.400	0.500
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.707	1.175	1.175	1.175	1.175	1.175	1.175	1.175	1.175	1.175	1.175	1.175	1.175	1.175
3	0.577	1.340	1.340	1.339	1.339	1.338	1.336	1.335	1.333	1.325	1.318	1.305	1.287	1.277
5	0.447	1.542	1.542	1.541	1.540	1.537	1.534	1.531	1.525	1.508	1.491	1.460	1.417	1.391
7	0.378	1.678	1.677	1.676	1.674	1.671	1.666	1.662	1.653	1.627	1.603	1.558	1.497	1.461
10	0.316	1.817	1.815	1.813	1.811	1.806	1.799	1.793	1.781	1.744	1.711	1.649	1.569	1.524
15	0.258	1.961	1.959	1.956	1.953	1.944	1.935	1.926	1.907	1.854	1.807	1.724	1.625	1.571
20	0.224	2.055	2.052	2.048	2.044	2.033	2.020	2.008	1.984	1.917	1.860	1.763	1.653	1.595
50	0.141	2.305	2.296	2.288	2.275	2.248	2.218	2.191	2.139	2.018	1.930	1.804	1.684	1.625
75	0.115	2.395	2.382	2.369	2.350	2.309	2.267	2.230	2.160	2.014	1.917	1.787	1.670	1.616
100	0.100	2.453	2.436	2.419	2.395	2.343	2.292	2.247	2.168	2.009	1.910	1.784	1.669	1.617
200	0.071	2.567	2.534	2.503	2.460	2.377	2.300	2.238	2.135	1.957	1.859	1.745	1.649	1.605
400	0.050	2.637	2.575	2.521	2.453	2.331	2.233	2.158	2.051	1.879	1.791	1.697	1.619	1.583
600	0.041	2.655	2.573	2.504	2.421	2.282	2.176	2.103	1.995	1.841	1.762	1.675	1.607	1.576
1000	0.032	2.645	2.534	2.451	2.354	2.211	2.109	2.037	1.940	1.798	1.716	1.643	1.601	1.570
2000	0.022	2.558	2.418	2.330	2.224	2.092	1.990	1.926	1.839	1.720	1.684	1.617	1.568	1.547
5000	0.014	2.443	2.268	2.203	2.121	1.989	1.856	1.794	1.746	1.665	1.616	1.603	1.522	1.540

Valores de RR (NO-EXCVOL)

N	$N^{-1/2}$	$Z= 0.003$	0.005	0.007	0.010	0.017	0.025	0.033	0.050	0.100	0.150	0.250	0.400	0.500
1	1.000	1.342	1.342	1.343	1.343	1.344	1.345	1.346	1.279	1.353	1.361	1.375	1.395	1.408
2	0.707	1.096	1.097	1.097	1.098	1.099	1.045	1.089	1.105	1.117	1.128	1.150	1.182	1.203
3	0.577	1.044	1.044	1.045	1.045	1.099	1.043	1.050	1.052	1.061	1.070	1.087	1.111	1.127
5	0.447	1.016	1.016	1.017	0.965	1.016	1.019	1.019	1.021	1.026	1.031	1.040	1.053	1.061
7	0.378	1.008	1.009	1.009	0.995	1.009	1.010	1.011	1.012	1.015	1.018	1.023	1.030	1.034
10	0.316	1.004	0.953	1.017	1.003	1.004	1.004	1.004	1.005	1.005	1.006	1.006	1.006	1.007
15	0.258	1.002	0.997	1.001	1.001	1.002	1.003	1.003	1.003	1.003	1.003	1.001	0.998	0.995
20	0.224	1.007	1.000	1.001	1.001	1.002	1.002	1.002	1.002	1.003	1.001	0.998	0.992	0.988
50	0.141	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.994	0.985	0.978
75	0.115	1.000	1.000	0.999	0.999	0.998	0.998	0.997	0.996	0.992	0.989	0.980	0.972	0.967
100	0.100	1.000	1.001	1.001	1.001	1.001	1.002	1.002	1.002	1.001	1.000	0.993	0.980	0.972
200	0.071	1.001	1.003	1.004	1.005	1.007	1.009	1.010	1.009	1.002	0.998	0.991	0.979	0.972
400	0.050	1.000	1.000	0.999	1.000	0.998	0.999	0.997	0.997	0.988	0.983	0.980	0.970	0.963
600	0.041	1.000	1.001	1.000	0.999	0.997	0.995	0.995	0.992	0.993	0.989	0.981	0.971	0.965
1000	0.032	0.998	0.997	0.999	0.997	1.002	1.004	1.003	1.003	0.996	0.975	0.970	0.982	0.972
2000	0.022	0.996	0.996	1.000	0.996	1.001	0.998	0.994	0.995	0.989	0.991	0.980	0.968	0.960
5000	0.014	1.007	1.010	1.010	1.014	1.010	0.994	0.985	0.989	0.985	0.983	0.978	0.951	0.959

Tabla IV.1: Resultados numéricos de la simulación Monte Carlo. GT , GI y $R_g/R_{g,BD}$ de cadenas vermiformes sin volumen excluido, como funciones de N y Z

IV.4. APLICACIONES

Valores de GT (EXCVOL)														
N	$N^{-1/2}$	$Z=0.003$	0.005	0.007	0.010	0.017	0.025	0.033	0.050	0.100	0.150	0.250	0.400	0.500
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.707	1.225	1.225	1.225	1.225	1.225	1.225	1.225	1.225	1.225	1.225	1.225	1.225	1.225
3	0.577	1.397	1.397	1.397	1.396	1.396	1.395	1.394	1.391	1.385	1.379	1.368	1.352	1.341
5	0.447	1.680	1.679	1.678	1.676	1.673	1.668	1.664	1.656	1.632	1.609	1.567	1.520	1.492
7	0.378	1.889	1.887	1.885	1.882	1.875	1.868	1.860	1.845	1.803	1.764	1.703	1.622	1.586
10	0.316	2.124	2.120	2.117	2.112	2.101	2.088	2.075	2.050	1.983	1.925	1.845	1.714	1.665
15	0.258	2.401	2.395	2.389	2.380	2.359	2.336	2.314	2.270	2.159	2.069	1.933	1.800	1.753
20	0.224	2.603	2.594	2.584	2.571	2.540	2.508	2.476	2.415	2.266	2.151	1.993	1.844	1.778
50	0.141	3.242	3.211	3.180	3.137	3.044	2.952	2.871	2.728	2.453	2.281	2.069	1.922	1.860
75	0.115	3.512	3.460	3.411	3.343	3.205	3.076	2.968	2.786	2.458	2.276	2.061	1.930	1.873
100	0.100	3.695	3.626	3.560	3.473	3.298	3.139	3.009	2.789	2.443	2.265	2.039	1.919	1.877
200	0.071	4.074	3.935	3.814	3.660	3.382	3.123	2.957	2.695	2.342	2.171	2.012	1.907	1.881
400	0.050	4.306	4.053	3.854	3.618	3.234	2.981	2.785	2.568	2.237	2.088	1.973	1.896	1.872
600	0.041	4.357	4.036	3.789	3.519	3.118	2.852	2.696	2.453	2.160	2.054	1.951	1.892	1.872
1000	0.032	4.307	3.909	3.647	3.329	2.932	2.692	2.526	2.330	2.100	2.013	1.920	1.887	1.872
2000	0.022	3.972	3.617	3.303	3.039	2.683	2.449	2.337	2.182	2.006	1.946	1.905	1.882	1.871
5000	0.014	3.540	3.103	2.855	2.627	2.394	2.215	2.144	2.062	1.914	1.918	1.892	1.878	1.871

Valores de GI (EXCVOL)														
N	$N^{-1/2}$	$Z=0.003$	0.005	0.007	0.010	0.017	0.025	0.033	0.050	0.100	0.150	0.250	0.	0.500
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.707	1.175	1.175	1.175	1.175	1.175	1.175	1.175	1.175	1.175	1.175	1.175	1.175	1.175
3	0.577	1.340	1.340	1.339	1.339	1.338	1.336	1.335	1.333	1.325	1.318	1.305	1.287	1.275
5	0.447	1.542	1.542	1.541	1.540	1.537	1.534	1.531	1.525	1.508	1.491	1.459	1.420	1.396
7	0.378	1.678	1.677	1.676	1.674	1.671	1.666	1.662	1.653	1.627	1.603	1.561	1.501	1.473
10	0.316	1.817	1.815	1.814	1.811	1.806	1.800	1.794	1.781	1.746	1.713	1.663	1.574	1.536
15	0.258	1.961	1.959	1.957	1.953	1.945	1.935	1.926	1.907	1.855	1.808	1.728	1.639	1.604
20	0.224	2.055	2.052	2.048	2.044	2.032	2.020	2.008	1.983	1.917	1.859	1.769	1.672	1.626
50	0.141	2.305	2.296	2.287	2.275	2.247	2.217	2.189	2.136	2.018	1.935	1.818	1.727	1.686
75	0.115	2.395	2.382	2.369	2.350	2.309	2.267	2.230	2.162	2.021	1.930	1.812	1.733	1.696
100	0.100	2.453	2.436	2.419	2.395	2.343	2.292	2.247	2.165	2.015	1.922	1.802	1.728	1.700
200	0.071	2.567	2.534	2.503	2.460	2.377	2.290	2.229	2.123	1.960	1.872	1.783	1.718	1.704
400	0.050	2.637	2.575	2.521	2.452	2.327	2.234	2.156	2.062	1.902	1.824	1.759	1.714	1.699
600	0.041	2.655	2.573	2.504	2.422	2.283	2.181	2.117	2.010	1.861	1.803	1.745	1.713	1.698
1000	0.032	2.646	2.538	2.459	2.356	2.211	2.111	2.039	1.944	1.827	1.781	1.727	1.708	1.696
2000	0.022	2.578	2.452	2.346	2.260	2.108	1.989	1.941	1.860	1.771	1.743	1.716	1.703	1.694
5000	0.014	2.426	2.267	2.171	2.075	1.978	1.870	1.840	1.818	1.707	1.723	1.706	1.698	1.693

Valores de α_g^2 (EXCVOL)														
N	$N^{-1/2}$	$Z=0.003$	0.005	0.007	0.010	0.017	0.025	0.033	0.050	0.100	0.150	0.250	0.400	0.500
5	0.447	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.997	0.997	1.009
100	0.100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.014	1.036	1.111	1.300	1.587
1000	0.032	0.998	1.009	1.016	1.015	1.021	1.022	1.032	1.029	1.124	1.245	1.429	-	-
5000	0.014	1.004	0.973	0.973	0.972	1.001	1.058	1.074	1.127	1.301	1.492	-	-	-

Values of Q (EXCVOL)													
$Z=0.010$	0.017	0.025	0.033	0.050	0.750	0.100	0.150	0.250	0.333	0.400	0.500	0.600	
0.095	0.126	0.147	0.169	0.194	0.253	0.284	0.334	0.419	0.479	0.524	0.598	0.682	

Tabla IV.2: Resultados numéricos de la simulación Monte Carlo. GT , GI , α_g^2 y Q_∞ ($N = \infty$), de cadenas vermiformes con volumen excluido, como funciones de N y Z

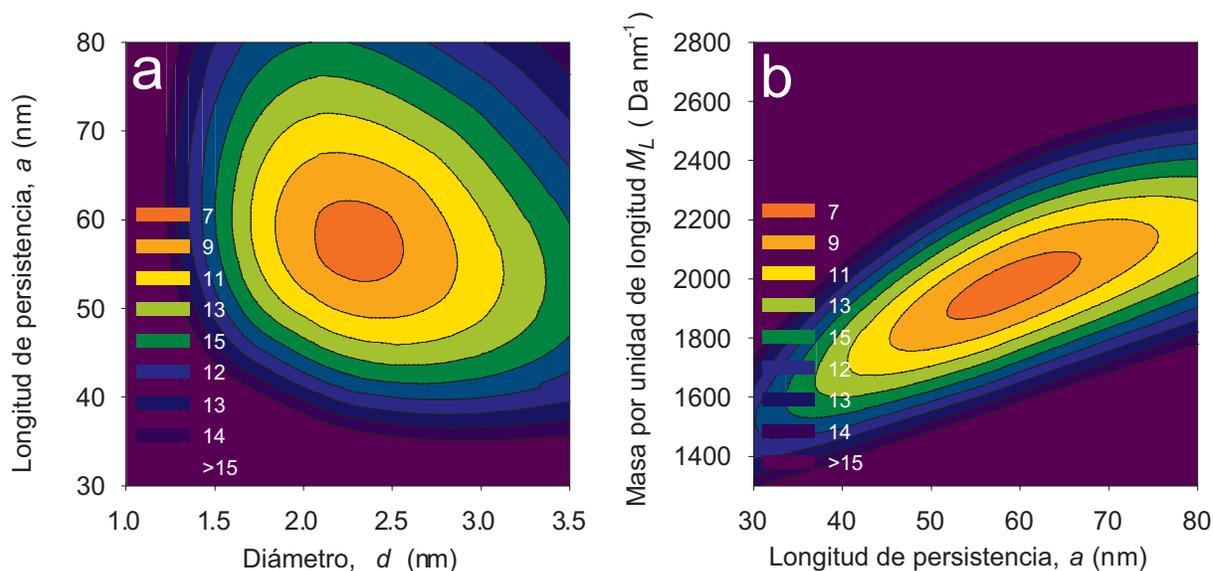


Figura IV.6: Gráficos de contorno del modelo de ADN de la función Δ respecto a dos parámetros estructurales para el ADN. El tercer parámetro es $M_L = 1950$ Da/nm en **a** y $d = 2.3$ nm en **b**.

IV.4.2. Ejemplos. ADN

El ADN en forma de doble hélice es el ejemplo paradigmático de molécula vermiforme. Existe una amplia bibliografía de datos experimentales, de la cual hemos recolectando 147 datos experimentales, que incluyen 27 radios de giro, 23 coeficientes de difusión, 56 coeficientes de sedimentación y 41 medidas de viscosidad intrínseca (ver Tablas IV.3, IV.4, IV.5 y IV.6), cubriendo un rango de longitudes desde 10 hasta 10^5 pares de base. Se han seleccionado medidas realizadas en presencia de fuerza iónica superior a 0.1 M, en cuyas condiciones se asume que la repulsión electroestática del ADN bicatenario queda neutralizada, por lo que el modelo vermiforme es perfectamente aplicable en este caso. En nuestro análisis, **Multi-HYDFIT** halló los tres parámetros barriendo amplios intervalos para los tres parámetros simultáneamente, dando como resultado los siguientes valores para dichos parámetros: $d = 2.3$ nm, $a = 56$ nm y $M_L = 1950$ Da/nm, con un valor de la función $\Delta = 6.6\%$, obteniéndose una excelente reproducción de los puntos experimentales, aunque se observa una ligera desviación en la viscosidad intrínseca (ver Figuras IV.7 y IV.8). **Multi-HYDFIT** realiza un ajuste simultáneo de las cuatro propiedades, por lo que

la calidad del ajuste puede variar de una propiedad a otra, particularmente en el caso de la viscosidad, muy sensible a los detalles estructurales.

La incertidumbre asociada a los parámetros hallados en el ajuste se pone en manifiesto en las gráficas de contorno de la función Δ respecto a dos parámetros, dejando el tercero fijado al valor hallado, en las Figuras IV.6a y b. Si elevamos ligeramente la tolerancia en el porcentaje de desviación al 7% ($\Delta \leq 0.07$), observamos que hay un rango de valores de los parámetros que consiguen reproducir los datos experimentales casi con la misma precisión, siendo el rango de $a = 52\text{-}64$ nm, de $d = 2.1\text{-}2.6$ nm, y de $M_L = 1850\text{-}2100$ nm, que podemos atribuir a la gran cantidad y diversidad de condiciones de medida y disoluciones empleadas en los datos experimentales incluidos en este análisis. Los parámetros hallados con un porcentaje de desviación igual o menor del 7% se muestran en la Tabla IV.7.

Si comparamos los parámetros d y a hallados en este estudio con los encontrados en la bibliografía, el diámetro d se encuentra dentro del rango 2.0 – 2.5 nm obtenidos experimentalmente por dispersión dinámica de luz (DLS), por anisotropía de polarización de fluorescencia o a partir de su estructura cristalográfica [111]. Respecto a la longitud de persistencia a , encontramos que el valor hallado en este trabajo ($a = 56$ nm) se encuentra muy próximo al obtenido por Hagerman *et al* [14], Elias *et al* [112], y Diekmann *et al* [113], y varía en torno a 460-700 nm según otros autores [111]. Por último, el parámetro M_L se encuentra muy próximo al calculado a partir del peso molecular promedio de un par de bases (660 Da) y su longitud promedio (0.34 nm), siendo este 1941.2 Da/nm.

Origen	Ref	M	D_t (cm ² /s)
pLH fragmentos	[114, 115]	1.53×10^6	4.56×10^{-8}
Linear ColE1	[116, 117]	4.34×10^6	2.03×10^{-8}
pLH fragmentos	[118]	1.50×10^6	4.56×10^{-8}
pLH fragmentos	[118]	6.70×10^5	7.15×10^{-8}
pLH fragmentos	[118]	5.00×10^5	9.05×10^{-8}
pLH fragmentos	[118]	2.40×10^5	1.58×10^{-7}
pLH fragmentos	[114]	1.53×10^6	4.42×10^{-8}
ColE1	[119, 114]	4.32×10^6	2.07×10^{-8}
phi29	[120]	1.15×10^7	1.00×10^{-8}
Mononucleosoma, fragmento	[121]	1.10×10^5	3.00×10^{-7}
Sintético	[122, 123]	5.28×10^3	1.53×10^{-6}
Sintético	[122, 123]	7.92×10^3	1.34×10^{-6}
Sintético	[122, 123]	1.32×10^4	1.09×10^{-6}
Sintético	[124]	6.60×10^3	1.40×10^{-6}
Sintético	[124]	7.92×10^3	1.45×10^{-6}
Sintético	[124]	7.92×10^3	1.22×10^{-6}
Sintético	[124]	9.24×10^3	1.17×10^{-6}
Sintético	[124]	1.06×10^4	1.11×10^{-6}
Sintético	[124]	7.92×10^3	1.23×10^{-6}
Sintético	[124]	9.24×10^3	1.19×10^{-6}
Sintético	[124]	1.58×10^4	9.50×10^{-7}
Sintético	[125]	1.32×10^4	1.07×10^{-6}
Sintético	[125]	1.32×10^4	1.02×10^{-6}
Sintético	[125]	1.32×10^4	1.00×10^{-6}
-	[126]	1.79×10^6	3.50×10^{-8}

Tabla IV.3: Coeficientes de difusión del ADN obtenidos de diferentes autores.

Origen	Ref.	M (Da)	R_g (cm)
pLH fragmento	[114, 115]	1.53×10^6	1.04×10^{-5}
Fago T7	[127]	2.60×10^7	7.10×10^{-5}
Fago T7	[127]	2.60×10^7	5.80×10^{-5}
ColE1	[128]	4.35×10^6	1.95×10^{-5}
ColE1	[116, 117]	4.34×10^6	1.86×10^{-5}
Fago T7	[129]	2.60×10^7	5.87×10^{-5}
Timo ternera, fragmento	[130]	2.81×10^5	3.46×10^{-6}
Timo ternera, fragmento	[130]	3.62×10^5	4.09×10^{-6}
Timo ternera, fragmento	[130]	4.67×10^5	4.87×10^{-6}
Timo ternera, fragmento	[130]	5.40×10^5	5.82×10^{-6}
Timo ternera, fragmento	[130]	6.57×10^5	6.43×10^{-6}
Timo ternera, fragmento	[130]	8.46×10^5	7.84×10^{-6}
Timo ternera, fragmento	[130]	1.27×10^6	1.05×10^{-5}
Fago T7	[131]	2.60×10^7	7.10×10^{-5}
Fago T7	[132]	2.60×10^7	5.20×10^{-5}
Fago T7	[133]	2.60×10^7	6.10×10^{-5}
Timo ternera, fragmentos	[134]	2.00×10^7	3.61×10^{-5}
Fago T7	[135]	2.60×10^7	7.20×10^{-5}
Calf-thymus, fragmento	[136]	3.05×10^5	4.59×10^{-6}
T7 y timo de ternera, fragmento	[137, 95, 138, 139, 135]	1.97×10^7	5.12×10^{-5}
T7 y timo de ternera, fragmento	[137, 95, 138, 139, 135]	1.56×10^7	4.05×10^{-5}
T7 y timo de ternera, fragmento	[137, 95, 138, 139, 135]	1.17×10^7	3.26×10^{-5}
T7 y timo de ternera, fragmento	[137, 95, 138, 139, 135]	5.82×10^6	2.40×10^{-5}
T7 y timo de ternera, fragmento	[137, 95, 138, 139, 135]	4.35×10^6	2.35×10^{-5}
T7 y timo de ternera, fragmento	[137, 95, 138, 139, 135]	3.45×10^6	1.86×10^{-5}
T7 y timo de ternera, fragmento	[137, 95, 138, 139, 135]	1.88×10^6	1.32×10^{-5}
T7 y timo de ternera, fragmento	[137, 95, 138, 139, 135]	6.22×10^5	7.01×10^{-6}
Fago T2	[140]	1.17×10^8	1.35×10^{-4}
1/2 fago T2	[140]	7.02×10^7	1.02×10^{-4}
1/4 fago T2	[140]	2.69×10^7	6.20×10^{-5}

Tabla IV.4: Radios de giro para el ADN obtenidos de diferentes autores.

Origen	Ref	M	s (S)
Fago PM2, fragmento	[141]	1.15×10^6	10.78
Fago PM2, fragmento	[141]	1.06×10^6	10.67
Fago PM2, fragmento	[141]	8.65×10^5	9.99
Fago PM2, fragmento	[141]	5.64×10^5	8.72
Fago PM2, fragmento	[141]	5.24×10^5	8.59
Fago PM2, fragmento	[141]	4.24×10^5	8.03
Fago PM2, fragmento	[141]	3.91×10^5	7.89
Fago PM2, fragmento	[141]	3.29×10^5	7.62
Fago PM2, fragmento	[141]	2.13×10^5	6.7
Fago PM2, fragmento	[141]	1.90×10^5	6.37
Fago PM2, fragmento	[141]	1.74×10^5	6.24
Fago PM2, fragmento	[141]	1.06×10^5	5.41
Fago PM2, fragmento	[141]	9.57×10^4	5.2
Fago PM2, fragmento	[141]	7.72×10^4	4.92
Fago PM2, fragmento	[141]	6.20×10^4	4.58
Fago PM2, fragmento	[141]	3.30×10^4	3.51
Linear ColE1	[116, 117]	4.34×10^6	16.6
Fago T7	[142]	2.60×10^7	31.8
Fago T5	[142]	7.90×10^7	51.78
Fago T4	[142]	1.13×10^8	62.06
Fago T7	[143]	2.60×10^7	31.3
Timo ternera, fragmento	[144]	2.43×10^5	7.2
Timo ternera, fragmento	[144]	1.91×10^5	6.3
Timo ternera, fragmento	[144]	1.14×10^5	5.7
Timo ternera, fragmento	[144]	9.00×10^4	5.4
Timo ternera, fragmento	[144]	8.80×10^4	5.4
Timo ternera, fragmento	[144]	7.30×10^4	4.7
Timo ternera, fragmento	[144]	4.50×10^4	3.9
Fago T2	[145, 146]	1.17×10^8	64.5
Fago T7	[145, 146]	2.60×10^7	33.7
1/2 T7, fragmento	[132, 147]	1.24×10^7	22.8
phiX 174RF	[132, 147]	3.58×10^6	14.3
Timo ternera, fragmento	[130]	2.81×10^5	7.79
Timo ternera, fragmento	[130]	3.62×10^5	8.34
Timo ternera, fragmento	[130]	4.67×10^5	8.89
Timo ternera, fragmento	[130]	5.40×10^5	9.36
Timo ternera, fragmento	[130]	6.57×10^5	9.8
Timo ternera, fragmento	[130]	8.46×10^5	10.38
Timo ternera, fragmento	[130]	1.27×10^6	11.73
Mononucleosoma, fragmento	[121]	1.10×10^5	5.33
Sintético	[124]	6.60×10^3	1.5
Sintético	[124]	7.92×10^3	1.77
Sintético	[124]	7.92×10^3	1.65
Sintético	[124]	9.24×10^3	1.85
Sintético	[124]	1.06×10^4	2.1
Fagos T2 y T7, fragmento	[146]	2.00×10^5	6.15
Fagos T2 y T7, fragmento	[146]	4.00×10^5	7.40
Fagos T2 y T7, fragmento	[146]	8.00×10^5	9.10
Fagos T2 y T7, fragmento	[146]	1.60×10^6	11.41
Fagos T2 y T7, fragmento	[146]	3.20×10^6	14.55
Fagos T2 y T7, fragmento	[146]	6.40×10^6	18.83
Fagos T2 y T7, fragmento	[146]	1.28×10^7	24.66
Fagos T2 y T7, fragmento	[146]	2.56×10^7	32.60
Fagos T2 y T7, fragmento	[146]	5.12×10^7	43.40
Fagos T2 y T7, fragmento	[146]	1.02×10^8	58.11
Fago T2	[140]	1.17×10^8	64.5
Half Fago T2	[140]	7.02×10^7	44
Quarter Fago T2	[140]	2.69×10^7	31.3

Tabla IV.5: Coeficientes de sedimentación del ADN obtenidos de diferentes autores.

Origen	Ref	M (Da)	$[\eta]$ (cm ³ /g)
Fago T7	[127]	2.60×10^7	1.06×10^4
Fago T7	[127]	2.60×10^7	1.10×10^4
Fago T2	[148, 149]	1.17×10^8	3.10×10^4
Fago T5	[149]	7.90×10^7	2.40×10^4
Fago T7	[149]	2.60×10^7	1.28×10^4
CT	[149]	7.00×10^6	4.60×10^3
Fago T7	[143]	2.60×10^7	1.25×10^4
Fago T2	[146, 145]	1.17×10^8	3.16×10^4
Fago T4	[142, 145]	1.13×10^8	3.13×10^4
Fago T5	[142, 145]	7.90×10^7	2.40×10^4
T5 sto	[142, 145]	6.23×10^7	2.26×10^4
Fago T7	[146, 145]	2.60×10^7	1.11×10^4
Fago T4 6x	[150, 147]	3.95×10^7	1.39×10^4
Fago T4 11x	[150, 147]	2.24×10^7	9.60×10^3
Timo ternera, fragmento	[130]	2.81×10^5	1.52×10^2
Timo ternera, fragmento	[130]	3.62×10^5	2.32×10^2
Timo ternera, fragmento	[130]	4.67×10^5	2.92×10^2
Timo ternera, fragmento	[130]	5.40×10^5	3.66×10^2
Timo ternera, fragmento	[130]	6.57×10^5	4.69×10^2
Timo ternera, fragmento	[130, 151]	8.46×10^5	5.82×10^2
Timo ternera, fragmento	[130, 151]	1.27×10^6	8.91×10^2
Fago T2	[152]	1.17×10^8	3.51×10^4
Fagos T2 y T7, fragmento	[146]	4.00×10^5	2.28×10^2
Fagos T2 y T7, fragmento	[146]	8.00×10^5	6.54×10^2
Fagos T2 y T7, fragmento	[146]	1.60×10^6	1.33×10^3
Fagos T2 y T7, fragmento	[146]	3.20×10^6	2.40×10^3
Fagos T2 y T7, fragmento	[146]	6.40×10^6	4.10×10^3
Fagos T2 y T7, fragmento	[146]	1.28×10^7	6.79×10^3
Fagos T2 y T7, fragmento	[146]	2.56×10^7	1.11×10^4
Fagos T2 y T7, fragmento	[146]	5.12×10^7	1.78×10^4
Fagos T2 y T7, fragmento	[146]	1.02×10^8	2.86×10^4
Timo ternera	[153]	7.80×10^6	4.73×10^3
Timo ternera, fragmento	[153]	2.06×10^5	1.39×10^2
Timo ternera, fragmento	[153]	3.99×10^5	2.34×10^2
Timo ternera, fragmento	[153]	3.62×10^5	2.22×10^2
Timo ternera, fragmento	[153]	2.55×10^5	1.57×10^2
Timo ternera, fragmento	[153]	1.65×10^5	8.90×10^1
Timo ternera, fragmento	[153]	7.61×10^4	3.00×10^1
Fago T2	[140]	1.17×10^8	2.77×10^4
1/2 fago T2	[140]	7.02×10^7	1.82×10^4
1/4 fago T2	[140]	2.69×10^7	1.13×10^4

Tabla IV.6: Viscosidades intrínsecas del ADN obtenidos de diferentes autores.

IV.4. APLICACIONES

d/nm	a/nm	$M_L/\text{Da nm}^{-1}$	100Δ	d/nm	a/nm	$M_L/\text{Da nm}^{-1}$	100Δ
2.1	52	1850	7.0	2.4	54	1950	6.9
2.1	54	1850	7.0	2.4	56	1900	7.0
2.1	54	1900	6.9	2.4	56	1950	6.7
2.1	56	1900	6.8	2.4	56	2000	6.8
2.1	56	1950	6.9	2.4	58	1950	6.7
2.1	58	1900	7.0	2.4	58	2000	6.6
2.1	58	1950	6.9	2.4	58	2050	6.9
2.1	60	1950	6.9	2.4	60	1950	6.8
2.2	52	1850	7.0	2.4	60	2000	6.6
2.2	52	1900	7.0	2.4	60	2050	6.7
2.2	54	1850	7.0	2.4	62	2000	6.7
2.2	54	1900	6.8	2.4	62	2050	6.6
2.2	54	1950	7.0	2.4	62	2100	6.9
2.2	56	1900	6.7	2.4	64	2000	6.9
2.2	56	1950	6.7	2.4	64	2050	6.7
2.2	58	1900	6.9	2.4	64	2100	6.8
2.2	58	1950	6.7	2.5	56	1950	6.9
2.2	58	2000	6.9	2.5	56	2000	6.9
2.2	60	1950	6.8	2.5	58	1950	6.9
2.2	60	2000	6.8	2.5	58	2000	6.7
2.2	62	1950	7.0	2.5	58	2050	6.9
2.2	62	2000	6.8	2.5	60	1950	7.0
2.2	64	2000	7.0	2.5	60	2000	6.6
2.3	52	1900	7.0	2.5	60	2050	6.6
2.3	54	1900	6.8	2.5	62	2000	6.7
2.3	54	1950	6.9	2.5	62	2050	6.6
2.3	56	1900	6.8	2.5	62	2100	6.8
2.3	56	1950	6.7	2.5	64	2000	7.0
2.3	56	2000	6.9	2.5	64	2050	6.7
2.3	58	1900	7.0	2.5	64	2100	6.7
2.3	58	1950	6.6	2.6	56	2000	7.0
2.3	58	2000	6.7	2.6	58	2000	6.8
2.3	60	1950	6.7	2.6	58	2050	6.9
2.3	60	2000	6.6	2.6	60	2000	6.8
2.3	60	2050	6.9	2.6	60	2050	6.7
2.3	62	1950	7.0	2.6	60	2100	7.0
2.3	62	2000	6.7	2.6	62	2000	6.9
2.3	62	2050	6.8	2.6	62	2050	6.7
2.3	64	2000	6.9	2.6	62	2100	6.8
2.3	64	2050	6.8	2.6	64	2050	6.7
2.4	54	1900	7.0	2.6	64	2100	6.7

Tabla IV.7: Conjunto de parámetros hallados por Multi-HYDFIT para el ADN, cuyo $100\Delta \leq 7\%$

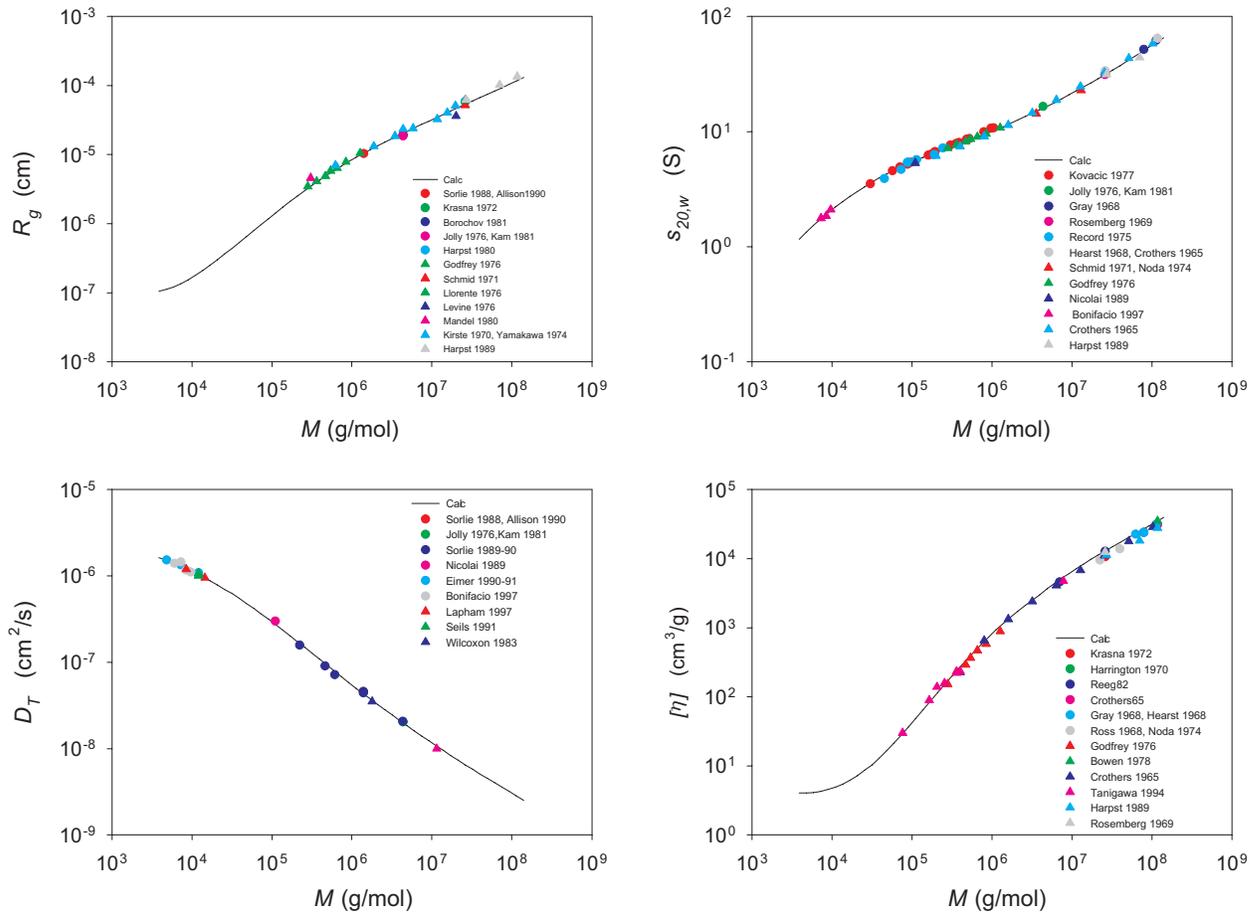


Figura IV.7: Comparación de las propiedades en disolución calculadas mediante Multi-HYDFIT del ADN (línea continua) respecto a las experimentales (puntos) del ADN.

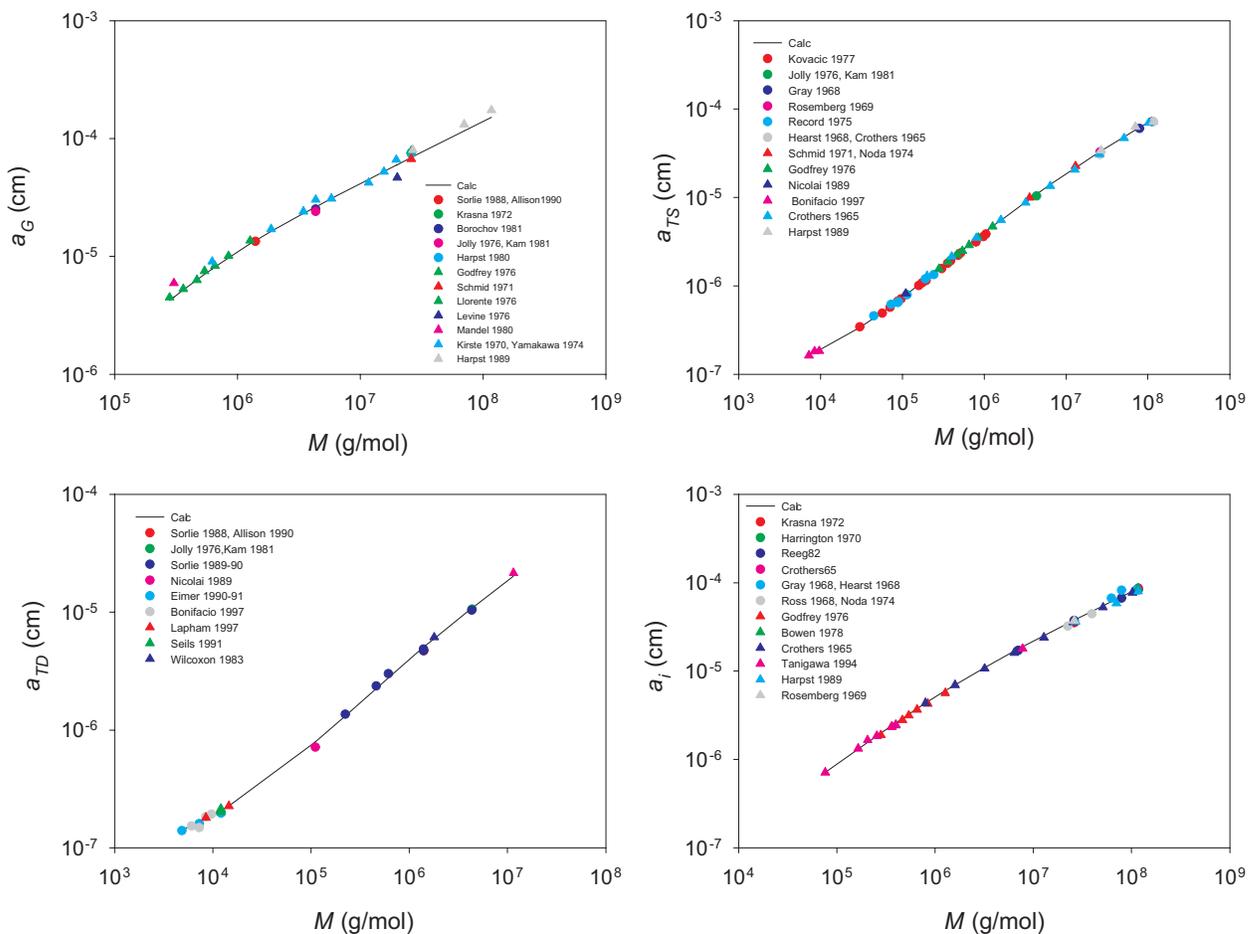


Figura IV.8: Comparación de los radios equivalentes calculados mediante Multi-HYDFIT del ADN (línea continua) respecto a los experimentales (puntos) del ADN.

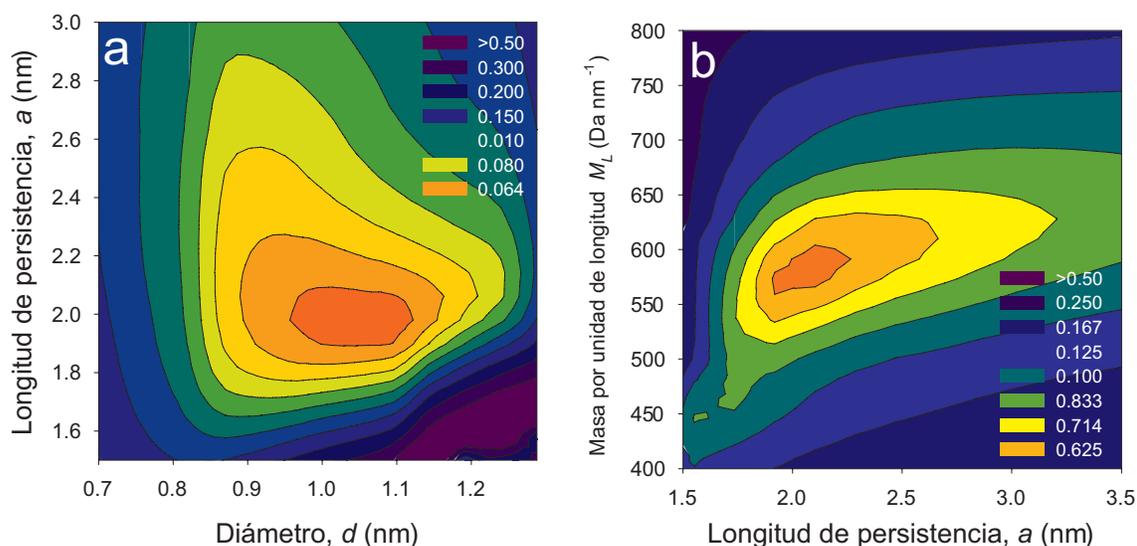


Figura IV.9: Gráficos de contorno del modelo de PSDFE de la función Δ respecto a dos parámetros estructurales. El tercer parámetro es $M_L = 600$ Da/nm en **a** y $d = 1.0$ nm en **b**.

IV.4.3. Otros polímeros

Poli(estireno-difeniletileno)

En colaboración con el grupo del Prof. George M. Pavlov (Universidades de S. Petersburgo y Jena), aplicamos por primera vez esta metodología con el fin de obtener los parámetros estructurales a partir de las propiedades en disolución de un copolímero vinílico flexible, sintetizado y caracterizado teórico-experimentalmente por nuestros colaboradores. Nuestra metodología se encontraba en una fase inicial, donde los efectos de volumen excluido todavía no estaban implementados. El trabajo fue publicado en el año 2009 [154].

Las propiedades en disolución de que disponemos son los coeficientes de sedimentación, difusión y viscosidad intrínseca, de cadenas poliméricas de bajo peso molecular. Pese a ser un polímero altamente flexible, debido a que son fragmentos relativamente cortos, es correcto y razonable aplicar el modelo vermiforme, y los efectos de volumen excluido pueden ser ignorados [93, 94, 155]. En la Tabla IV.8 podemos observar un resumen de las propiedades en disolución obtenidas experimentalmente, a partir de las cuales se realizó el ajuste global y simultáneo, empleando los resultados obtenidos de nuestra simulación

M	$s_{20,w}$ S	D_t $\times 10^8 \text{ cm}^2/\text{s}$	$[\eta]$ cm^3/g
25200	0.93	57.75	13.1
15700	0.76	76.22	10.4
11900	0.64	82.67	8.3
6000	0.48	128.70	5.9
3700	0.36	155.96	5
2600	0.29	171.21	3.7
1700	0.25	207.85	3

Tabla IV.8: Resumen de las propiedades en disolución obtenidas experimentalmente para las diversas fracciones del polímero.

MC y, alternativamente, las teorías de Yamakawa-Fujii (YF), ambos incorporados en el programa `Multi-HYDFIT`.

Los resultados del ajuste dan como resultado $d = 1.1 \text{ nm}$, $a = 2.1 \text{ nm}$ y $M_L = 600 \text{ Da/nm}$, con un valor de la función $\Delta = 5.1 \%$. En la Figura IV.9 se muestran los gráficos de contorno de la función Δ respecto a dos parámetros, dejando el tercero fijado al valor hallado. Estas predicciones derivadas del tratamiento MC para la masa por unidad de longitud, M_L y diámetro d , están muy en consonancia con las estimaciones basadas en la estructura química y en el volumen específico [154], mientras que la longitud de persistencia, a , es alrededor de dos veces la del poliestireno, que puede ser explicado por la presencia de los anillos fenílicos que impiden la rotación interna. Si realizamos el ajuste empleando el tratamiento de Yamakawa-Fujii (YF), obtenemos los parámetros $d = 0.53 \text{ nm}$, $a = 1.8 \text{ nm}$ y $M_L = 360 \text{ Da/nm}$, con un valor de la función $\Delta = 12.8 \%$, notablemente peor que el ajuste que se obtiene con nuestro tratamiento. La implementación de los resultados de MC elimina los posibles defectos de la interacción hidrodinámica promediada en el tratamiento de Yamakawa-Fujii, e incluye una mejor descripción del efecto del diámetro hidrodinámico, muy relevante en el caso de cadenas poliméricas cortas. Las predicciones son mucho mejores empleando el método MC, además de obtener un valor de la función Δ mucho más bajo (5.1 % frente a 12.8 %), mejorándose notablemente la bondad del ajuste. Las propiedades y radios equivalentes predichos junto a los experimentales se muestran en la Figura IV.10.

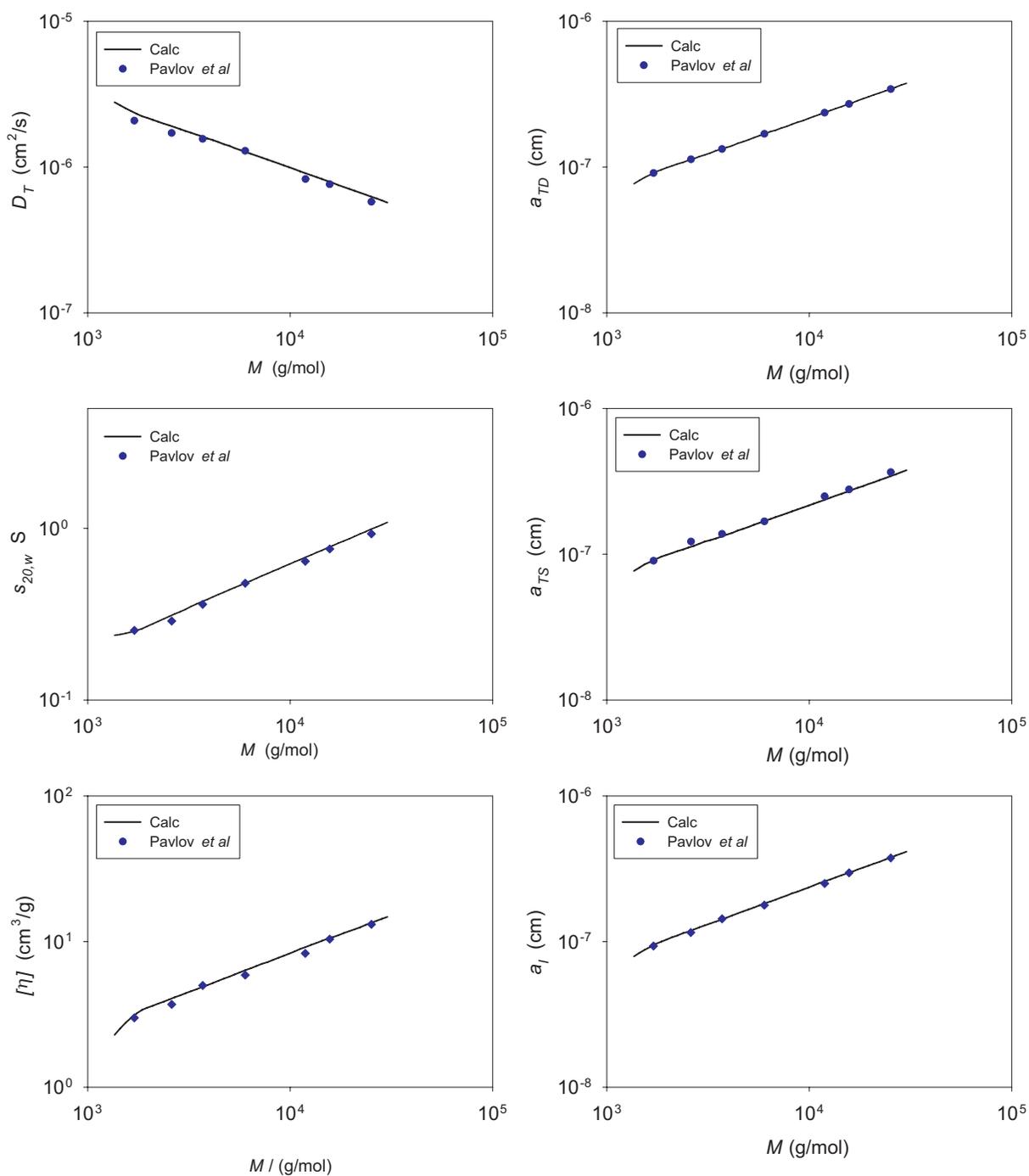


Figura IV.10: Comparación de las propiedades en disolución y radios equivalentes calculados (línea continua) respecto a las experimentales (puntos) del poli(estireno-difeniletileno).

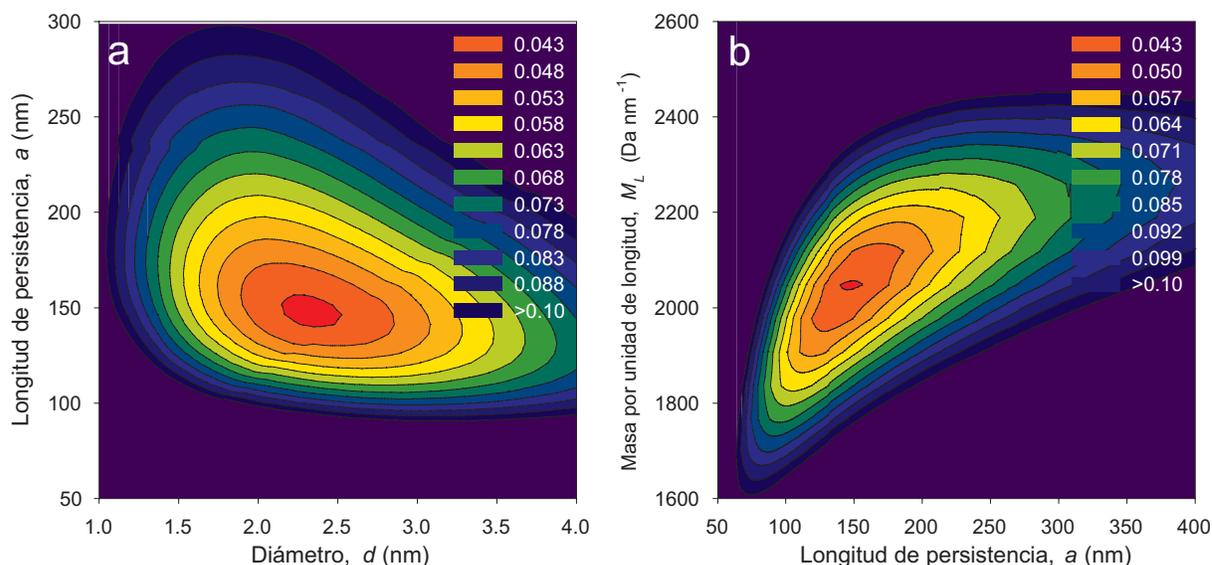


Figura IV.11: Gráficos de contorno del modelo de esquizofilano de la función Δ respecto a dos parámetros estructurales. El tercer parámetro es $M_L = 2050 \text{ Da/nm}^{-1}$ en **a** y $d = 2.3 \text{ nm}$ en **b**.

Esquizofilano

Polímero de glucosa (β -glucano) de origen fúngico, que en disolución forma una triple hélice. Se trata de un polímero muy rígido, con especial importancia en la industria fármaco-alimentaria por poseer propiedades inmunoestimulantes. Los datos experimentales fueron obtenidos de los trabajos de Kashiwagi *et al* [156] y Yanaki *et al* [157] (ver Tabla IV.9). Los parámetros hallados en este estudio son $d = 2.3 \text{ nm}$, $a = 150 \text{ nm}$ y $M_L = 2050 \text{ Da/nm}$, consistentes con los estimados a partir de su estructura química de triple hélice, además de coincidir con los hallados empleando las teorías de YF por Kashiwagi *et al*, cuyos valores varían en torno a $d = 2.2 - 2.6 \text{ nm}$, $a = 150 - 200 \text{ nm}$ y $M_L = 2000 - 2200 \text{ Da/nm}$. La concordancia en ese caso entre nuestros resultados y los obtenidos con YF puede atribuirse a que los datos experimentales caen en este caso bien dentro del rango de validez de la teoría YF.

M g/mol	$s_{20,w}$ S	R_g $\times 10^8$ cm	$[\eta]$ cm ³ /g	Ref.
5680000	-	3330	12000	[156]
4000000	-	2510	7450	[156]
2850000	-	2050	4380	[156]
1510000	-	1470	1920	[156]
620000	-	723	580	[156]
461000	-	572	455	[156]
413000	-	510	330	[156]
353000	-	448	295	[156]
259000	-	347	166	[156]
170000	-	221	79.1	[156]
107000	-	146	38.4	[156]
5700000	15	-	12000	[157]
4000000	13	-	7190	[157]
2480000	11.2	-	4370	[157]
1610000	10.3	-	2500	[157]
1000000	9.12	-	1540	[157]
724000	8.74	-	962	[157]
429000	8.01	-	478	[157]
320000	7.36	-	301	[157]
196000	6.57	-	132	[157]
139000	6.09	-	73.6	[157]
96000	5.25	-	40.1	[157]

Tabla IV.9: Resumen de las propiedades en disolución obtenidas experimentalmente para las diversas fracciones de esquizofilano. Las medidas fueron tomadas en agua 0.01 N NaOH a 25° C.

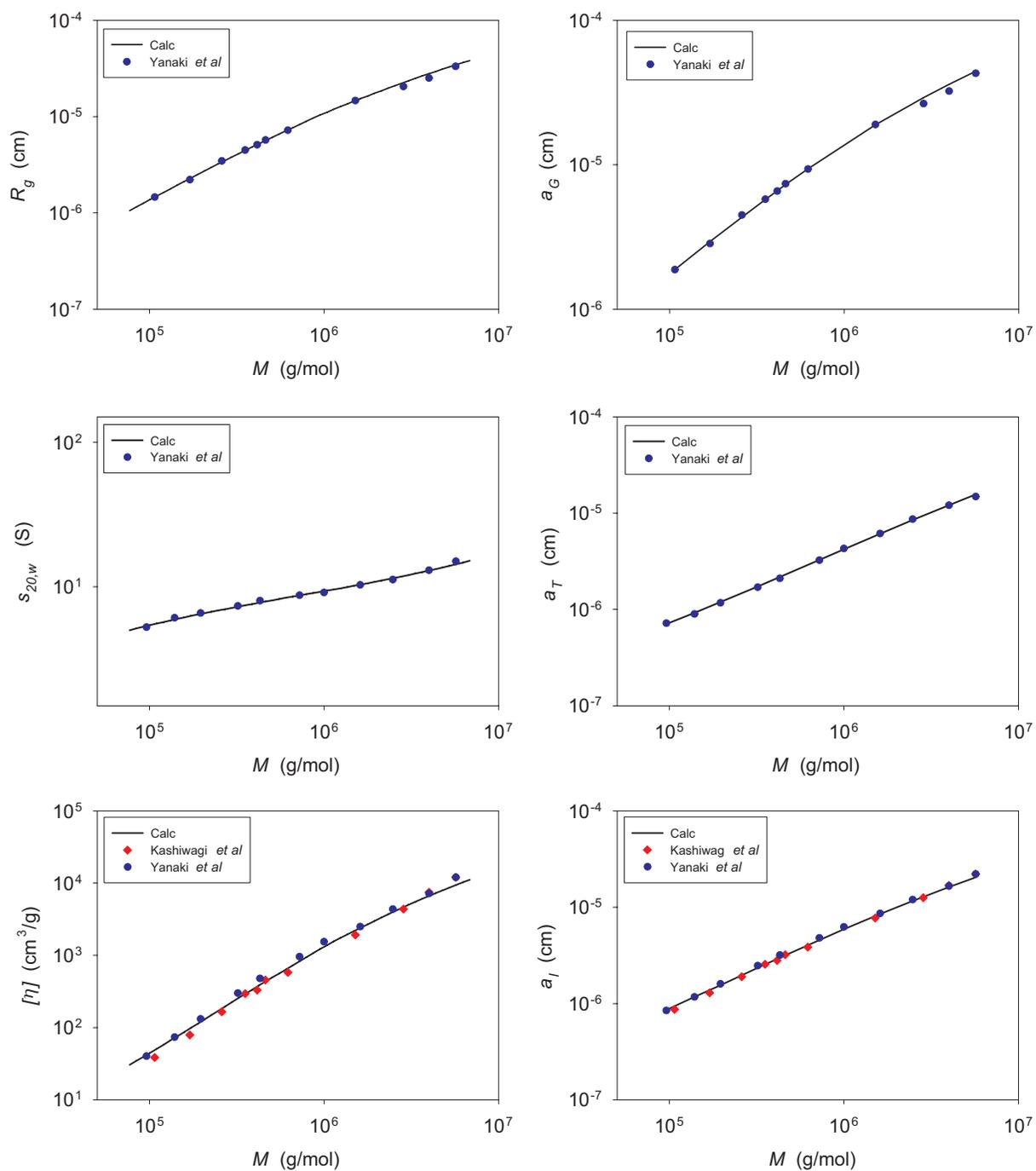


Figura IV.12: Comparación de las propiedades en disolución y radios equivalentes calculados (línea continua) respecto a las experimentales (puntos) del esquizofilano.

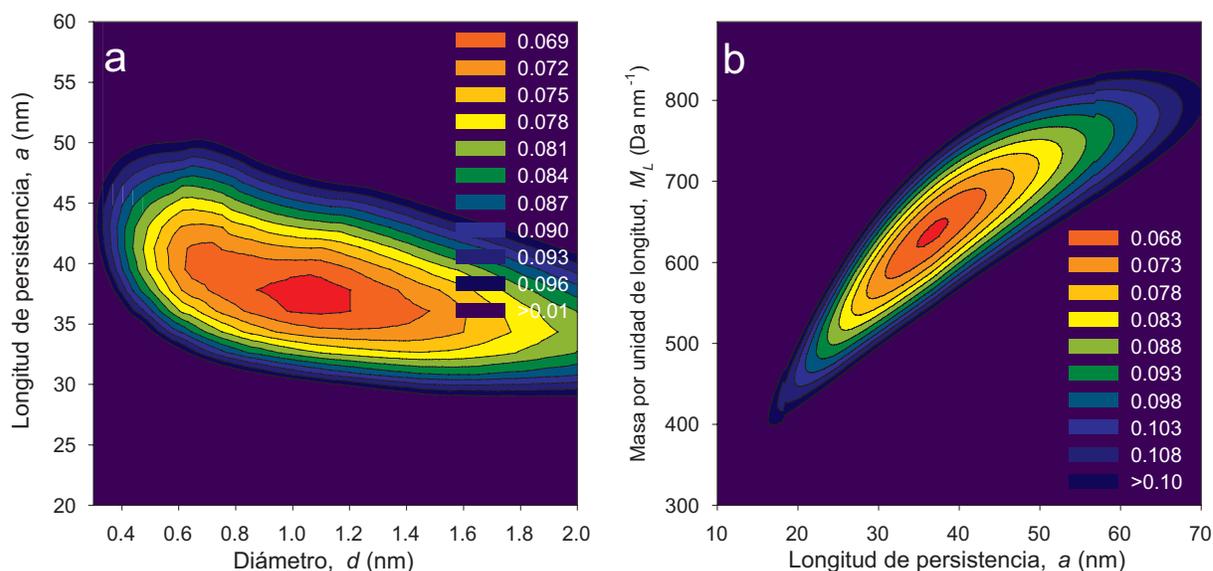


Figura IV.13: Gráficos de contorno del modelo de PHIC de la función Δ respecto a dos parámetros estructurales. El tercer parámetro es $M_L = 64.5 \text{ Da}/\text{\AA}$ en **a** y $d = 10.0 \text{ \AA}$ en **b**.

Poli(*n*-hexil isocianato)

El poli(*n*-hexil isocianato) (PHIC) es un polímero considerado cristal líquido a altas concentraciones y capaz de formar geles termorreversibles a concentraciones moderadas. Ha sido objeto de numerosos estudios en las últimas tres décadas por su habilidad para adoptar conformaciones extendidas debido a su peculiar estructura química. Los datos experimentales fueron obtenidos de los trabajos de Murakami *et al* [158] y Rubingh *et al* [159] (ver Tabla IV.10). Los parámetros hallados en este estudio son $d = 1.0 \text{ nm}$, $a = 37 \text{ nm}$ y $M_L = 645 \text{ Da}/\text{nm}$. Murakami *et al* [158] realizan una estimación del diámetro $d \approx 1.3 \text{ nm}$, basándose en la estructura cristalográfica del poli(*n*-butil isocianato), y este autor, junto a Rubingh *et al* [159] obtienen los parámetros $a = 42 \text{ nm}$ y $M_L = 640\text{-}715 \text{ Da}/\text{nm}$ empleando las teorías de YF.

M g/mol	$s_{20,w}$ S	R_g $\times 10^8$ cm	$[\eta]$ cm^3/g	Ref.
68000	6.71	226.94	251	[158]
81400	7.34	261.92	310	[158]
112000	7.87	327.11	426	[158]
132000	8.1	379.47	550	[158]
188000	8.81	472.23	770	[158]
255000	9.57	593.3	1110	[158]
311000	10.2	683.37	1400	[158]
387000	11.4	776.53	1720	[158]
505000	13.1	914.33	2250	[158]
656000	13.5	1063	2900	[158]
947000	17.1	1284.5	4040	[158]
1170000	18	1442.2	4700	[158]
1490000	18.2	1634	6040	[158]
1690000	20	1769.2	6360	[158]
2410000	22	2151.7	8600	[158]
3180000	25.3	2508	10500	[158]
3340000	26	2582.6	10900	[158]
4630000	30.8	3162.3	14100	[158]
7240000	36.4	4171.3	20000	[158]
100000	-	380	360	[159]
320000	-	920	465	[159]
460000	-	990	1120	[159]
520000	-	1100	1580	[159]
700000	-	1200	2820	[159]
2300000	-	2000	9500	[159]
5500000	-	3400	-	[159]

Tabla IV.10: Resumen de las propiedades en disolución obtenidas experimentalmente para las diversas fracciones de polihexilisocianato. Las medidas fueron tomadas en hexano a 25° C.

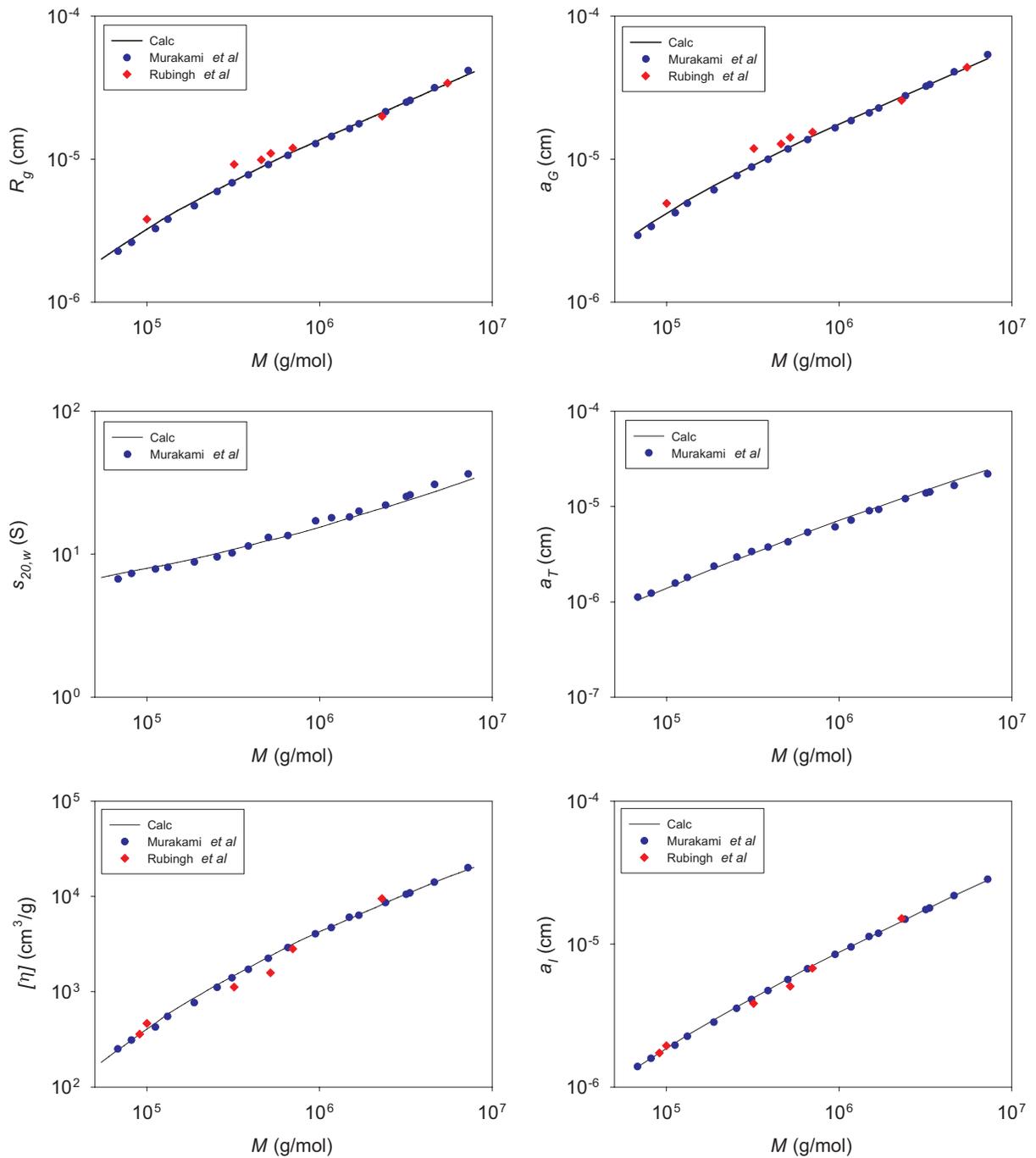


Figura IV.14: Comparación de las propiedades en disolución y radios equivalentes calculados (línea continua) respecto a las experimentales (puntos) del poli(*n*-hexil isocianato),

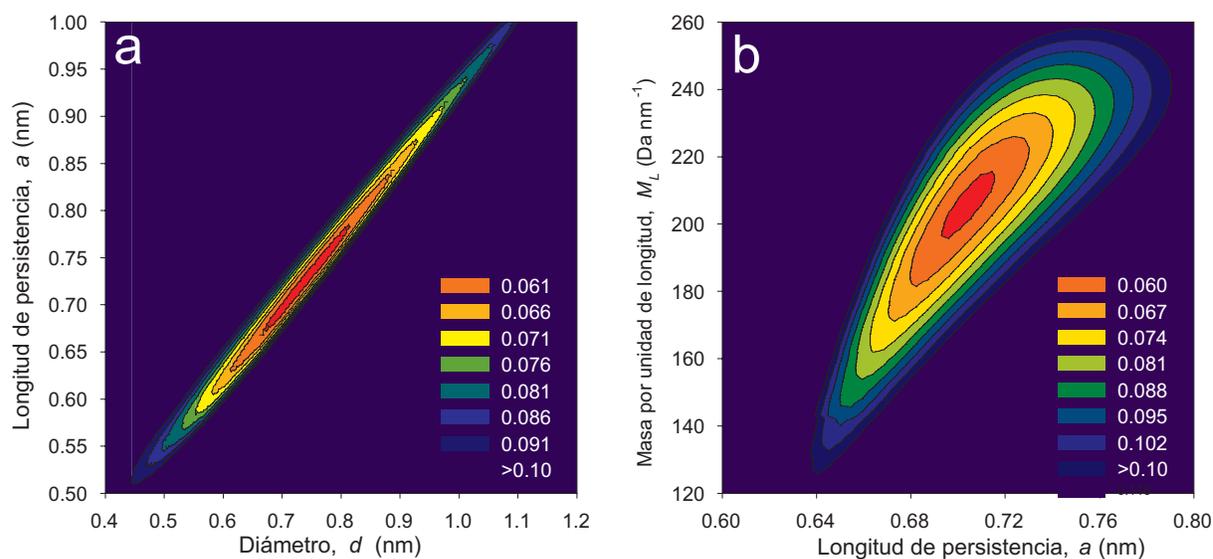


Figura IV.15: Gráficos de contorno del modelo de PIB de la función Δ respecto a dos parámetros estructurales. El tercer parámetro es $M_L = 205$ Da/nm en **a** y $d = 0.69$ nm en **b**.

Poli(isobutileno)

El poli(isobutileno) (PIB) es un caucho sintético, o elastómero. Es un polímero vinílico, de estructura muy similar al polietileno y al polipropileno excepto en que uno de los carbonos está sustituido por dos grupos metilo. Se trata de un polímero muy flexible, en el límite de aplicabilidad del modelo vermiforme. Los datos experimentales fueron obtenidos de los trabajos de Abe *et al* [160], Osa *et al* [161], Yamada *et al* [162] y Fetters *et al* [163] (ver Tabla IV.11). Los parámetros obtenidos mediante Multi-HYDFIT son: $d = 0.69$ nm, $a = 0.71$ nm y $M_L = 205$ Da/nm, valores próximos a los estimados a partir de su estructura química: $d \approx 0.5$ nm y $M_L = 216$ Da/nm (en cadena completamente extendida, en conformación todo *trans*) [162].

M g/mol	D_t $\times 10^8$ cm ² /s	R_g $\times 10^8$ cm	$[\eta]$ cm ³ /g	Refs.
112	-	-	0.3	[160]
224	-	-	1.6	[160]
641	6.5	-	3.2	[162, 160]
792	7.7	-	3.4	[162, 160]
1010	8.9	741.0	3.8	[162, 161, 160]
1250	10.0	-	4.1	[162, 160]
1470	11.0	604.0	4.5	[162, 161, 160]
1810	12.0	-	4.9	[162, 160]
8730	23.0	-	11.9	[162, 160]
16600	45.0	163.0	17.4	[162, 161, 160]
27900	57.0	-	23.2	[162, 160]
48500	-	95.0	33.0	[161, 160]
85500	-	-	47.1	[160]
130000	-	55.0	61.5	[161, 160]
422000	262.0	28.8	134.0	[161, 160, 161]
634000	324.0	23.3	180.0	[161, 160, 161]
819000	389.0	19.9	217.0	[161, 160, 161]
1760000	606.0	12.9	378.0	[161, 160, 161]
64000	85.9	83.5	-	[163]
13500	-	-	15.0	[163]
27000	-	-	21.3	[163]
84000	-	-	41.9	[163]
133000	-	-	60.0	[163]
167000	-	-	71.3	[163]
330000	224.0	33.4	107.0	[163]
449000	-	-	137.0	[163]
746000	360.0	21.1	206.0	[163]
1540000	538.0	14.4	310.0	[163]

Tabla IV.11: Resumen de las propiedades en disolución obtenidas experimentalmente para las diversas fracciones de poliisobutileno. Las medidas fueron tomadas en n-heptano a 25° C.

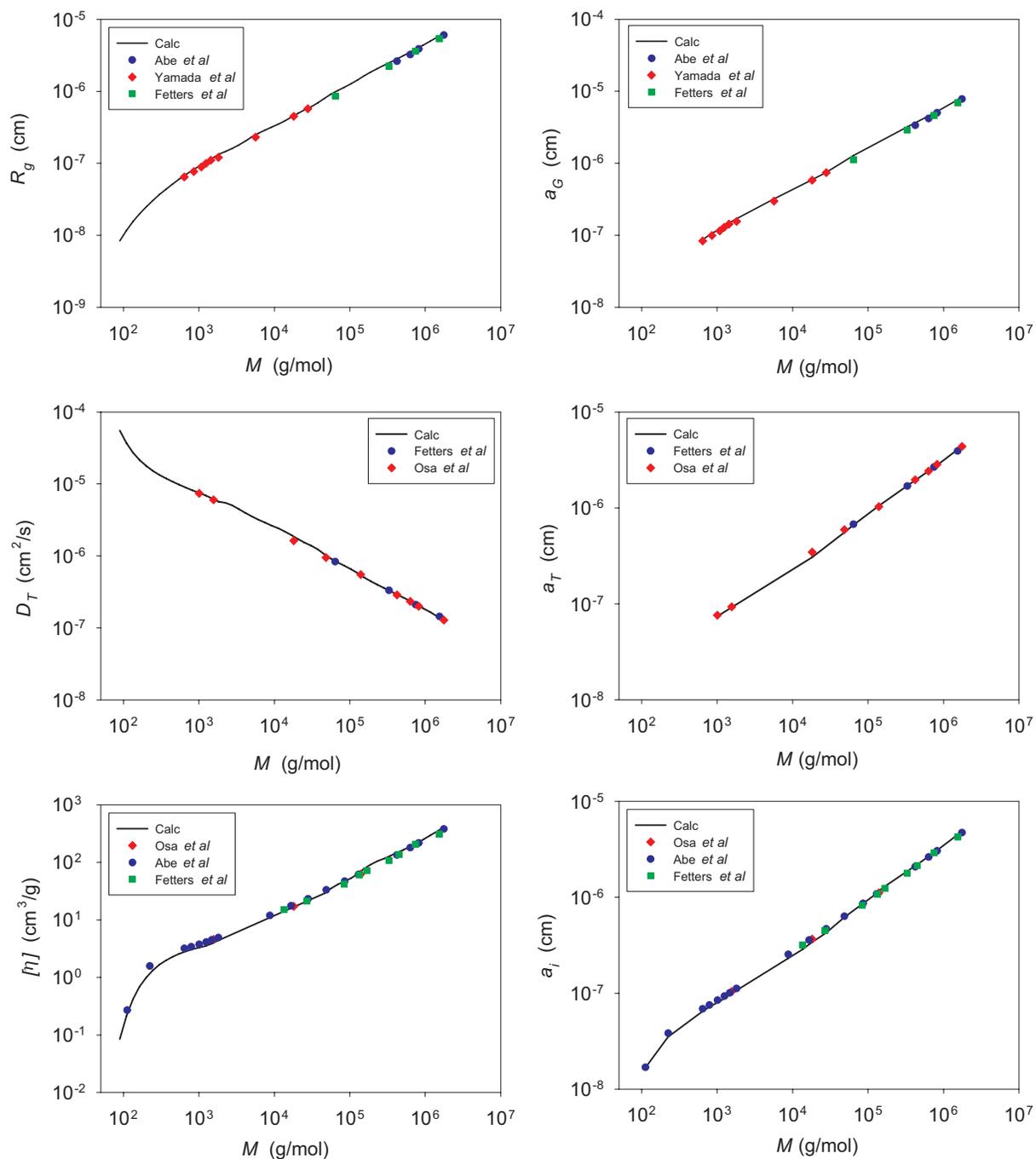


Figura IV.16: Comparación de las propiedades en disolución y radios equivalentes calculados (línea continua) respecto a las experimentales (puntos) del poli(isobutileno).

	M / Da	a / nm	$M_L / \text{Da nm}^{-1}$	d / nm	100Δ	L/a	n_{tot}
ADN	$6.0 \times 10^3 - 1.2 \times 10^8$	56	1950	2.3	6.6	0.6 - 1100	147
Poli(estireno-difeniletileno)	$1.7 \times 10^3 - 2.52 \times 10^4$	2.1	600	1.1	5.1	1.35 - 20	21
Esquizofilano	$9.6 \times 10^4 - 5.7 \times 10^6$	150	2050	2.3	4.2	0.3 - 18	44
Poli(<i>n</i> -hexil isocianato)	$6.8 \times 10^4 - 5.5 \times 10^6$	37	645	1.0	6.7	2.8 - 230	70
Poli(isobutileno)	$1.12 \times 10^2 - 5.5 \times 10^6$	0.71	205	0.69	6.2	0.8 - 38500	57

Tabla IV.12: Tabla resumen de los resultados de todos los polímeros.

IV.4.4. Comparación con otras metodologías

A continuación, con el objetivo de validar nuestro método, se compararán los resultados obtenidos con los de otras metodologías en la literatura científica. Además de contrastar los resultados de MC con y sin volumen excluido, se compararán ambos frente a las teorías clásicas de Yamakawa-Fujii [93, 95], Benoit-Doty [102], a las ecuaciones de García-Molina [101] y frente a una descripción hidrodinámica del modelo vermiforme de reciente publicación, propuesta por Mansfield y Douglas [164]. El método consiste en seleccionar unos parámetros del modelo elegido y comparar las propiedades, en forma de radios equivalentes, obtenidas por estos tratamientos, frente a los obtenidos por MC con VE. Mediante división simple obtenemos el *ratio* $a_{Xi} = a_{Xi}(A)/a_{Xi}(B)$, siendo *A* el tratamientos a estudio, y *B* en este caso siempre MC-VE. Los resultados se muestran en la Figura IV.17.

Empleando los parámetros del ADN ($d = 2.3 \text{ nm}$, $a = 50 \text{ nm}$ y $M_L = 1950 \text{ Da/nm}$) (Fig.IV.17, columna izquierda) encontramos muy buena concordancia entre nuestros cálculos de Monte Carlo VE y la aproximación propuesta por Mansfield y Douglas (puntos amarillos), dentro del rango de validez cubierto por este último (70-6000 pares de bases), que es bastante más restringido que el nuestro; el porcentaje de diferencia en el radio de giro y propiedades hidrodinámicas entre los dos métodos es alrededor del 7%, lo que confirma la validez de ambas aproximaciones para el caso del ADN. Respecto a la teoría clásica de Yamakawa-Fujii (puntos rojos excepto a_G , que se calcula por la ec. de Benoit-Doty), encontramos que la concordancia es bastante aceptable, incluso para pesos moleculares elevados. Las desviaciones son mínimas en fragmentos cortos de ADN, pero aumentan significativamente a medida que aumenta la longitud de la cadena, hasta llegar al 15% de diferencia en el caso del radio hidrodinámico traslacional en fragmentos de ADN de alto peso molecular. La aproximación que conlleva la teoría YF provoca una sobreestimación del radio hidrodinámico, mientras que la ausencia de volumen excluido produce el efecto

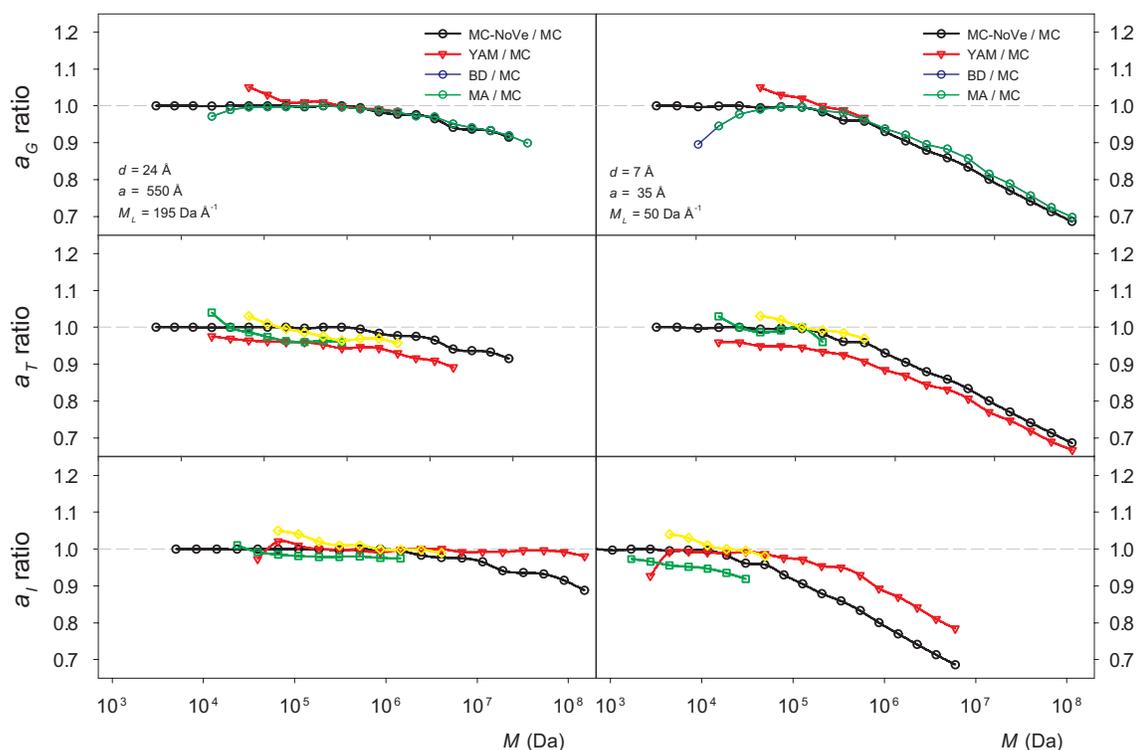


Figura IV.17: Comparación de metodologías para el cálculo de propiedades en disolución de polímeros vermiformes, con los parámetros del ADN (columna izquierda) y los de un polímero flexible (columna derecha). Códigos **MC**: Monte Carlo con VE, **MC-noVE**: Monte Carlo sin VE, **BD**: ecuación Benoit-Doty, **YF**: teoría de Yamakawa-Fujii, **GM**: ecuaciones de García-Molina, **MD**: Mansfield-Douglas.

contrario, por lo que se compensan. Además, con estos parámetros, el efecto de volumen excluido solo se aprecia a pesos moleculares elevados, como puede verse al comparar los resultados de MC-noVE respecto a los de MC-VE (puntos negros).

Si empleamos parámetros de un polímero con un alto grado de flexibilidad ($d = 7$ nm, $a = 35$ nm y $M_L = 50$ Da/nm) (Fig. IV.17, columna derecha) se observan notables desviaciones. En este caso, el volumen excluido tiene un efecto importante. Sin embargo, las propiedades calculadas coinciden muy bien con las obtenidas de la aproximación propuesta por Mansfield y Douglas, aunque como ocurría en el caso anterior, con un rango de aplicación mucho más restringido. Respecto a la teoría de YF, esta no contempla el efecto de VE, desviándose más a medida que aumenta M , acercándose mucho más a los resultados del cálculo MC-noVE (puntos negros), que obviamente, tampoco contempla VE. Nótese la gran concordancia con las ecuaciones de García-Molina en ambos casos.

IV.4.5. Resumen y conclusiones

La anterior comparación de teorías ha puesto de manifiesto que algunas funcionan bien en ciertos rangos de los parámetros del modelo vermiforme, pero fallan en otros; ninguna cubre el modelo en su totalidad.

En la tabla IV.12 se compilan los resultados obtenidos con nuestros tratamientos para los parámetros estructurales de las muy diversas macromoléculas vermiformes, mediante nuestro análisis `Multi-HYDFIT`, de los datos experimentales frente a los resultados de cálculos de nuestro procedimiento de Monte Carlo.

La aludida diversidad se refiere a varios aspectos. Por ejemplo al origen, natural o biológico de los polímeros, con su consiguiente diversidad de composición química y estructura local. Además, el rango de pesos moleculares, desde 10^2 hasta 10^8 . Dentro de su propio rango de pesos moleculares, cada macromolécula cubre un rango de la variabilidad conformacional del modelo vermiforme; así, hemos cubierto L/a desde 0.6 hasta 38500, con polímeros extremadamente flexibles como el poli(isobutileno) o extremadamente rígidos, como el esquizofilano, y con rangos de la relación M_L desde aprox. 200 hasta 2000. En todos los casos, el error típico para conjuntos tan amplios de propiedades y muestras esta entre el 4% y el 7%, bastante próximo a la incertidumbre propia de los datos experimentales.

Este conjunto de logros es imposible con cualquiera de los tratamientos anteriores del modelo vermiforme, por lo que cabe esperar que nuestro tratamiento, tanto en cuanto al de cálculo, como al de análisis, sean profusamente utilizados en el futuro.

Capítulo V

Proteínas

V.1. Introducción

Las proteínas son macromoléculas de gran importancia biológica, puesto que la práctica totalidad de las funciones de las células de todo ser vivo dependen de la acción de este tipo de moléculas. Por un lado, forman parte de la estructura básica de los tejidos y, por otro, desempeñan funciones metabólicas y reguladoras.

Durante muchos años, la aplicación principal de las propiedades hidrodinámicas en proteínas (y otras macromoléculas), ha sido la determinación del tamaño y forma de estas, mediante modelos hidrodinámicos simples, como el elipsoide de revolución. Años más tarde, la cristalografía de rayos X y la NMR multidimensional se convirtieron en las técnicas más empleadas para la obtención de información estructural en el estudio de proteínas (típicamente globulares) y otras macromoléculas, que hicieron posible la determinación con gran detalle de estructuras de macromoléculas y complejos macromoleculares. No obstante, las propiedades en disolución han continuado siendo una alternativa o complemento de las técnicas de alta resolución, especialmente para la validación de estas estructuras.

Paralelamente, en las últimas décadas, se han producido avances en la determinación estructural en macromoléculas donde las técnicas de alta resolución no son aplicables. Concretamente, proteínas en cuyo estado nativo, funcional, la cadena polipeptídica en buena parte de su extensión está desplegada. Son las llamadas proteínas “intrínsecamente desordenadas”. Se estima que el 10 % de todas las proteínas están completamente desordenadas, y el 40 % de las proteínas de eucariotas contienen al menos un segmento largo desestructurado (más de 50 residuos) [165]. Muchas de estas proteínas (o segmentos) se

encuentran desordenados bajo ciertas condiciones, y se estructuran en el momento que interaccionan con otra biomolécula, cambian de medio (membranas) o sufren una modificación post-traducciona, entre otras causas [166]. Este tipo de proteínas cuestiona una de las piedras angulares en la biología molecular, el paradigma estructura-función, donde una función específica está determinada por una única estructura tridimensional rígida [167].

Algunas técnicas indirectas, como el dicroísmo circular, la espectroscopia de infrarrojo (FTIR) o la fluorescencia únicamente nos revelan la pérdida de la estructura secundaria de forma cualitativa. Otras técnicas de gran utilidad como la NMR en sus modalidades de relajación (parámetros T_1 , T_2 , NOE (efecto Overhauser Nuclear) o RDC (acoplamiento dipolar residual), cuyos fundamentos están regidos por la hidrodinámica, han sido desarrolladas y aplicadas al estudio de este tipo de proteínas. Pero la técnica más empleada actualmente para la caracterización de este tipo de proteínas es sin duda la dispersión de rayos X de bajo ángulo (SAXS). Junto al desarrollo de complejos algoritmos de modelado molecular, esta técnica es capaz de mostrar el tamaño y la forma global de la proteína, siendo necesaria la validación posterior de estas estructuras de baja resolución mediante el cálculo de sus propiedades en disolución y la comparación con las medidas experimentales. Estos modelos de baja resolución (denominados de “grano grueso” o “coarse-grained”) suelen estar compuestos por esferas que no representan átomos (pseudo-átomos) o bien esferas centradas en la posición de algún átomo (típicamente carbono α) del esqueleto peptídico.

La necesidad de teorías y metodologías de cálculo y simulación, que correlacionen estructura y propiedades se hace más evidente con todos estos avances. En este capítulo de la Tesis abordaremos primero el estudio de proteínas globulares rígidas, y trataremos de extender la metodología del programa HYDROPRO a estructuras de baja resolución, además de realizar una parametrización rigurosa de este tipo de modelos, y en segundo lugar, continuando con la simulación numérica de macromoléculas en disolución, abordaremos el estudio de proteínas “intrínsecamente desordenadas”, donde desarrollaremos una metodología computacional adecuada para el estudio de este tipo de proteínas, desde el punto de vista de la conformación e hidrodinámica.

V.2. Cálculo hidrodinámico de proteínas rígidas

Actualmente, la base teórica y la implementación computacional del modelado de esferas, en el caso de (bio)macromoléculas rígidas, están muy desarrollados. Como se ha comentado en la Introducción de esta Tesis, para el cálculo de propiedades en disolución de proteínas globulares rígidas, se desarrollaron los programas HYDROPRO e HYDRONMR, que trabajan a partir de coordenadas atómicas.

El procedimiento de HYDROPRO consiste en reemplazar cada átomo (excepto H) por una esfera en el denominado modelo primario. El radio de esta esfera se estima en $\approx 3 \text{ \AA}$, por lo que cada esfera solapa considerablemente con sus vecinas, además de que la mayoría se encuentran en el interior de la estructura, no expuestas al disolvente, por lo que la mayoría de los elementos del modelo no ejercen ninguna influencia sobre las propiedades calculadas. Este hecho aumenta innecesariamente el coste computacional, que crece exponencialmente con el número de esferas (N^3), además de hacerse impracticable para proteínas compuestas por un número elevado de átomos. Para evitar estos problemas, en lugar de calcular las propiedades a partir del modelo primario, la versión anterior de HYDROPRO empleaba la estrategia del modelo de concha [12]. Si recordamos, se sustituyen las esferas del modelo primario (atómico) por un conjunto de mini-esferas tangentes en la superficie de la estructura, cuyo número va aumentando a medida que el radio (σ) de estas mini-esferas disminuye, realizándose sucesivos cálculos y obteniendo las propiedades por extrapolación a $\sigma = 0$, desde una concha compuesta por 400 esferas hasta un máximo de 2000, independientemente del tamaño de la estructura. Se ha demostrado que con esta metodología es posible reproducir con notable exactitud las propiedades en disolución experimentales de proteínas globulares rígidas y fragmentos cortos de ADN [12], y se ha usado ampliamente para el análisis de los coeficientes hidrodinámicos y otras propiedades en disolución.

Como comentamos anteriormente, existe una tendencia actual a emplear modelos reducidos o de “grano grueso” (“*coarse-grained*”). Concretamente, nos ocuparemos del caso en los que los elementos representan no átomos, sino residuos completos. Así, hemos abordado el cálculo de propiedades en disolución de proteínas a partir de modelos en los que cada elemento representa un residuo, extendiendo de esta manera la metodología a estructuras con menor resolución, donde únicamente se resuelven las coordenadas de los carbonos α ($C\alpha$). Hemos desarrollado una nueva versión del programa HYDROPRO en la que la estructura de partida puede ser una a nivel de residuo, conteniendo, simplemente las

coordenadas de los $C\alpha$. Además, otra novedad es que este cálculo con modelos reducidos puede hacerse con la metodología hidrodinámica del modelo de concha, o directamente a partir del modelo primario, con una esfera por residuo.

La parametrización del radio hidrodinámico para los modelos atómicos y reducidos la llevaremos a cabo empleando la metodología de radios equivalentes (ver sección II.1.3) tomando como referencia un amplio conjunto de estructuras atómicas de proteínas globulares (obtenidas por cristalografía de rayos X o NMR), que pertenecen a proteínas cuyas propiedades en disolución se han medido experimentalmente. Estas estructuras se muestran en la Tabla V.1. Las propiedades consideradas en este estudio son: R_g , D_t , $[\eta]$ y D_r , y se compararán en forma de radios equivalentes (a_G , a_T , a_I y a_R respectivamente).

Como hemos indicado, el nuevo programa HYDROPRO presenta tres modos de cálculo de propiedades en disolución, que son:

- a) A partir de la estructura con detalle atómico, generamos el modelo primario compuesto por esferas con radio r_{EA} (una esfera por cada átomo diferente al hidrógeno), lo sustituimos por el modelo de concha y obtenemos las propiedades por extrapolación a $\sigma = 0$.
- b) A partir del modelo simplificado (CG), generamos el modelo primario $C\alpha$ con un radio $r_{C\alpha-shell}$, (una esfera por cada $C\alpha$) lo sustituimos por el modelo de concha, y obtenemos las propiedades por extrapolación a $\sigma = 0$.
- c) A partir del modelo simplificado (CG), generamos el modelo primario $C\alpha$ con un radio $r_{C\alpha}$ (una esfera por cada $C\alpha$), y calculamos directamente las propiedades en disolución.

Una vez obtenidas las propiedades en disolución por estos tres métodos, se compararán las propiedades calculadas con las experimentales, en forma de radios equivalentes mediante la función Δ , ya empleada anteriormente:

$$\Delta^2 = \frac{1}{N_{prop}} \sum_{N_{prop}} \frac{1}{N_{prot}} \sum_{N_{prot}} \left[\frac{a_X(cal) - a_X(exp)}{a_X(exp)} \right]^2 \quad (V.1)$$

donde N_{prop} es el número de propiedades consideradas en este estudio, y N_{prot} es el número total de estructuras de proteínas con propiedades en disolución conocidas. De nuevo, $100\sqrt{\Delta^2}$ es el porcentaje de desviación de los radios equivalentes calculados respecto a los experimentales ($< \% \text{ dif.}>$, Tabla V.2), y sirve como medida global de la bondad del ajuste sobre las cuatro propiedades, para cada radio hidrodinámico probado (r_{EA} , $r_{C\alpha_{(Shell)}}$ o $r_{C\alpha}$). Una vez obtenido el radio óptimo global, se comprueba la desviación en ese radio para cada una de las propiedades ($\% \text{ dif. } a_G, a_T, a_I \text{ y } a_R$, ver Tabla V.2). Se empleó el programa HYDRPRO versión 10, en una máquina Windows 7 x64 provista de procesador Xeon T5500 2.8 GHz y 12 GB de RAM.

V.2.1. Proteínas de tamaño bajo-medio

Se analizaron un total de 132 datos experimentales, 54 coeficientes de difusión, 47 viscosidades intrínsecas, 8 radios de giro y 23 coeficientes de difusión rotacional, para un total de 69 estructuras analizadas (ver Tabla V.1), provenientes de tres compilaciones de datos realizados por autores diferentes: García de la Torre *et al* [12, 168] (set 1), S. Aragón *et al* [169] (set 2) y M. Rocco *et al* [170] (set 3). Se excluyeron del análisis las estructuras 8CAT (set 1), 1BGL (set 2) y 1A4V (set 3) por no disponer de la especie multimérica correspondiente. Además, en el set 2, 1HHO fue reemplazada por 1GZX, mientras que en el set 3, 5LDH y 1HCO fueron reemplazados por 6LDH y 2DN3 respectivamente.

Resultados

Los resultados globales se muestran en la Figura V.1 y Tabla V.2, mientras que los resultados del análisis parcial de los sets individuales (por autor) se muestran en las Figuras V.2, V.3 y V.4, y Tablas V.3, V.4 y V.5. Además de hallar un único radio óptimo donde las propiedades consideradas experimentan el mínimo error en su conjunto, se realizó un análisis de las propiedades individuales, con el fin de hallar nuevamente un radio óptimo, pero en este caso para cada propiedad por separado. Los resultados de este análisis parcial se muestra en la Tabla V.6.

Tras analizar conjuntamente los tres sets de datos, observamos que los errores obtenidos del modelo atómico y del modelo simplificado de $C\alpha$ (casos a y b) apenas difieren, siendo ambos prácticamente iguales (4.4 frente al 4.6 %), por lo que la concordancia con los datos experimentales apenas se ve afectada al reducir el detalle del modelo. Sin embargo, se debe emplear un radio elevado para el $C\alpha$ (4.8 Å frente al radio del modelo

atómico, 2.8 Å). En el caso de los cálculos prescindiendo del modelo de concha (caso c), observamos que el error se mantiene en el orden de los casos anteriores (4.8 %) pero debe emplearse un radio para el $C\alpha$ todavía más elevado que con los casos anteriores (6.1 Å), obteniéndose un mínimo en ese radio para las propiedades calculadas (ver Fig. V.1c).

Si analizamos los datos provenientes de los tres autores por separado, encontramos resultados similares (similar radio óptimo y error asociado) entre los sets 1,2 y 3, en los casos de los modelo atómicos (caso a), $C\alpha - shell$ (caso b) y $C\alpha$ (caso c), teniendo en cuenta que el set 2 no incluye medidas de R_g ni D_r , y el set 3 no incluye medidas de R_g .

En resumen, concluimos que en el caso de proteínas globulares de tamaño bajo-medio, podemos prescindir del modelo atómico detallado, siendo recomendable la metodología del modelo de concha, empleando radios adecuados para el $C\alpha$ (5.1 Å). Sin embargo, si nos vemos forzados a prescindir del modelo de concha, se pueden calcular las propiedades, de nuevo empleando radios adecuados para el $C\alpha$ (6.1 Å), con un error muy similar respecto a las propiedades experimentales.

PDB	M (Da)	a_T	a_I	a_G	a_R	Ref.	Set
4PTI	6158	16.64			15.69	[12]	1
1RBX	13700	20.06	19.28	19.11		[12]	1
6LYZ	14320	19.69	18.96	18.90	20.01	[12]	1
1MBO	17190	19.87	20.69	21.30	21.28	[12]	1
2CGA	25660	23.08	22.50	23.37		[12]	1
1BEB	36730	27.44			27.79	[12]	1
1OVA	43500	26.96	28.90			[12]	1
1CTS	97938	37.00	39.44	37.57		[12]	1
4GPD	142868	42.92	42.76	41.44		[12]	1
6LDH	145169	42.50	44.39	44.80		[12]	1
1ADO	156000	48.23	48.10			[12]	1
2MIN	220000	53.65				[12]	1
1SVN	26700				22.88	[168]	1
1BVG	21580				23.36	[168]	1
1LKI	19100				24.32	[168]	1
6I1B	17400				22.88	[168]	1
1STN	15510				23.42	[168]	1
1HWA	14320				20.01	[168]	1
1WRT	11890				28.15	[168]	1
1BTA	10140				19.26	[168]	1
1UBQ	8540				17.34	[168]	1
1CLB	8430				16.79	[168]	1
2BCA	8430				17.01	[168]	1
1EGL	8150				18.16	[168]	1
1PIT	6160				16.20	[168]	1
1ZNF	2930				13.23	[168]	1
1HRC	12400	18.50	17.53			[169]	2
7RSA	13700	20.09	19.47			[169]	2
1HFX	14200	20.27	19.32			[169]	2
2CDS	14300	19.69	18.58			[169]	2
1MBO	17200	20.54	20.70			[169]	2
1AVU	20100	21.90	20.74			[169]	2
1TPO	23300	23.08	22.54			[169]	2
1TGN	24000	22.17	22.42			[169]	2
4CHA	25200	21.04	22.89			[169]	2
2CGA	25700	23.25	22.55			[169]	2
2CAB	28800	24.14	24.50			[169]	2
1ZAG	32600	33.02	29.57			[169]	2
4PEP	34500	25.67	26.02			[169]	2
1J6Z	43000	28.54	29.33			[169]	2
6TAA	52500	29.12	30.17			[169]	2
1AO6	66500	35.18	35.60			[169]	2
1OVT	76000	36.37	35.78			[169]	2
1LFG	77100	38.32	36.57			[169]	2
2SOD	32500	25.95	25.72			[169]	2
1BEB	36700	29.40	28.07			[169]	2
4CHA	50400	29.60	32.19			[169]	2
1GKB	51400	34.61	32.21			[169]	2
8TIM	53200	31.75	31.63			[169]	2
2AAI	61500	35.77	30.68			[169]	2
1HHO	63200	31.65	30.28			[169]	2
1ALK	94600	37.65	37.08			[169]	2
1CTS	98000	37.00	39.45			[169]	2
1FAJ	117300	37.65	42.06			[169]	2
1ADO	157100	47.17	45.61			[169]	2
4BLC	235700	52.34	52.63			[169]	2
1BGL	465800	68.57	65.36			[169]	2
8RAT	13682	18.50	19.27		19.81	[170]	3
1A4V	15793	19.69			21.51	[170]	3
1DWR	17521	20.06			21.30	[170]	3
2CGA	25666	22.59	23.55			[170]	3
1BEB	36608	27.34			28.19	[170]	3
1OVA	43157	27.76	30.14		27.23	[170]	3
1HCO	64557	31.10				[170]	3
1GZX	64573	29.77	31.87		32.46	[170]	3
1CTS	97838	37.00	39.43			[170]	3
2GD1	143540	42.92				[170]	3
1GD1	146431	40.49	43.11			[170]	3
5LDH	148636	42.41				[170]	3
1ADO	157122		44.89			[170]	3
4BLC	235762	52.34	52.63			[170]	3
N		54	47	8	23		

Tabla V.1: Conjunto de proteínas de los tres conjuntos considerados en este estudio, junto a sus propiedades conocidas.

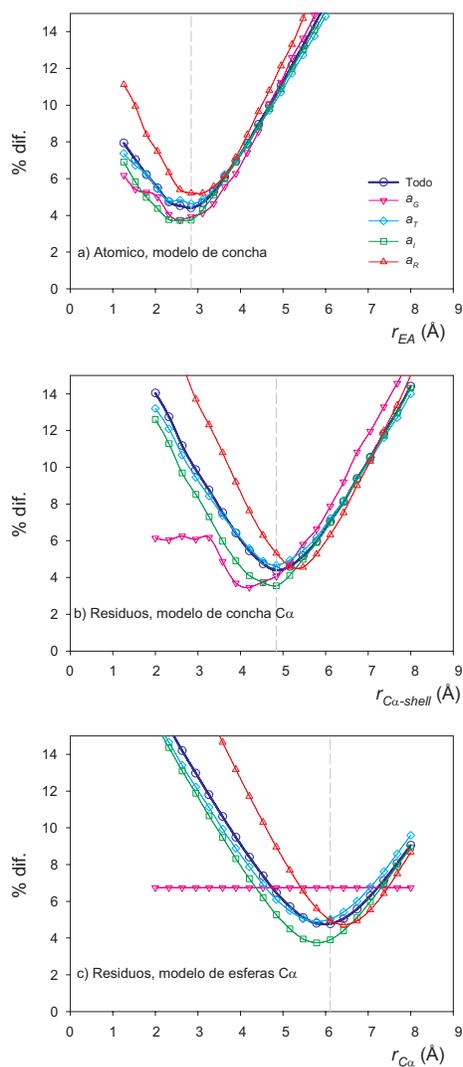


Figura V.1: Resultados para el set completo de estructuras, de la optimización del radio hidrodinámico r_{EA} del modelo de concha atómico (a), del radio $r_{C\alpha-shell}$ para el modelo de concha reducido (b), y del radio $r_{C\alpha}$ del modelo de esferas reducido (c).

Caso	Modelo	r óptimo (Å)	$\langle \% \text{ dif.} \rangle$	$\% \text{ dif. } a_G$	$\% \text{ dif. } a_T$	$\% \text{ dif. } a_I$	$\% \text{ dif. } a_R$
a	Atómico	2.84	4.41	3.92	4.63	3.75	5.21
b	$C\alpha(shell)$	4.84	4.39	4.05	4.66	3.54	5.32
c	$C\alpha$	6.11	4.78	6.74	5.03	3.90	4.96

Tabla V.2: Resultados de las minimizaciones para el set completo de estructuras, considerando todas las propiedades conjuntamente.

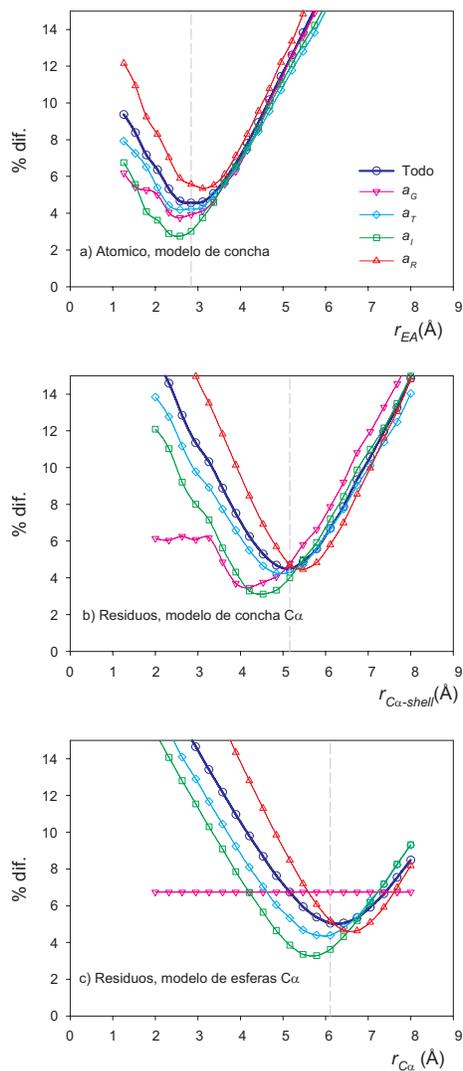


Figura V.2: Resultados para el set 1 de estructuras, de la optimización del radio hidrodinámico r_{EA} del modelo de concha atómico (a), del radio $r_{C\alpha-shell}$ para el modelo de concha reducido (b), y del radio $r_{C\alpha}$ del modelo de esferas reducido (c).

Caso	Modelo	r óptimo (Å)	$\langle \% \text{ dif.} \rangle$	$\% \text{ dif. } a_G$	$\% \text{ dif. } a_T$	$\% \text{ dif. } a_I$	$\% \text{ dif. } a_R$
a	Atómico	2.84	4.56	3.92	4.23	3.01	5.58
b	$C\alpha_{(shell)}$	5.16	4.48	4.77	4.29	4.01	4.71
c	$C\alpha$	6.11	5.05	6.74	4.38	3.62	5.19

Tabla V.3: Resultados de las minimizaciones para el set 1 de estructuras, considerando todas las propiedades conjuntamente.

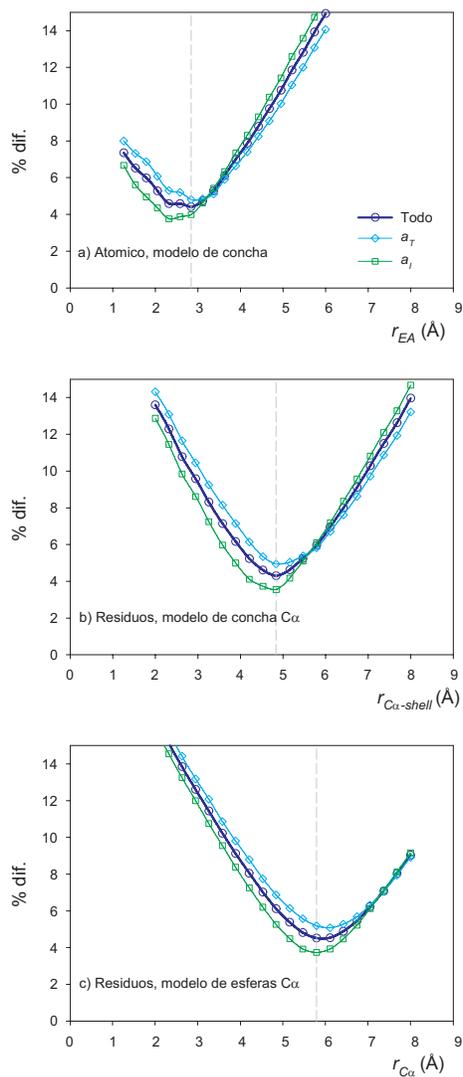


Figura V.3: Resultados para el set 2 de estructuras, de la optimización del radio hidrodinámico r_{EA} del modelo de concha atómico (a), del radio $r_{C\alpha-shell}$ para el modelo de concha reducido (b), y del radio $r_{C\alpha}$ del modelo de esferas reducido (c).

Caso	Modelo	r óptimo (Å)	< % dif. >	% dif. a_G	% dif. a_T	% dif. a_I	% dif. a_R
a	Atómico	2.84	4.41	-	4.8	3.98	-
b	$C\alpha_{(shell)}$	4.84	4.31	-	4.95	3.55	-
c	$C\alpha$	5.79	4.52	-	5.18	3.73	-

Tabla V.4: Resultados de las minimizaciones para el set 2 de estructuras, considerando todas las propiedades conjuntamente.

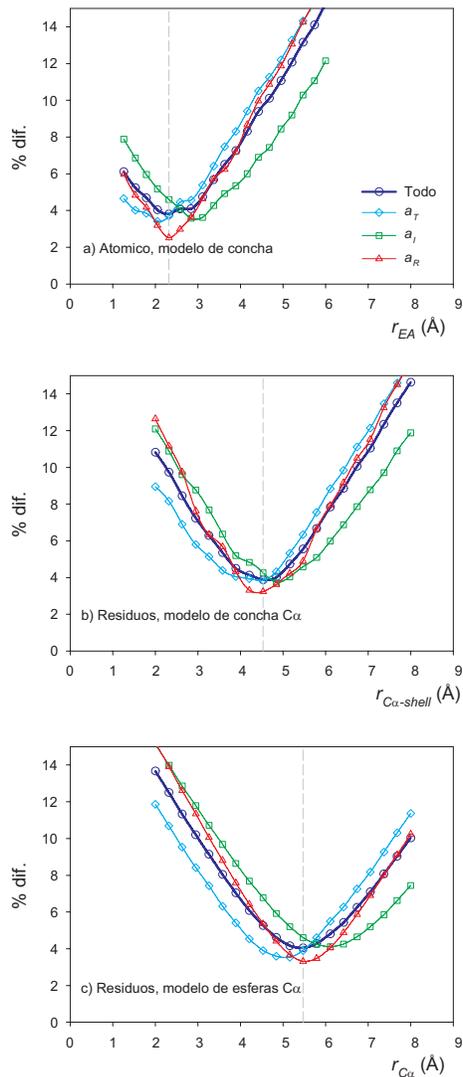


Figura V.4: Resultados para el set 3 de estructuras, de la optimización del radio hidrodinámico r_{EA} del modelo de concha atómico (a), del radio $r_{C\alpha-shell}$ para el modelo de concha reducido (b), y del radio $r_{C\alpha}$ del modelo de esferas reducido (c).

Caso	Modelo	r óptimo (Å)	$\langle \% \text{ dif.} \rangle$	$\% \text{ dif. } a_G$	$\% \text{ dif. } a_T$	$\% \text{ dif. } a_I$	$\% \text{ dif. } a_R$
a	Atómico	2.32	3.82	-	3.69	4.60	2.52
b	$C\alpha_{(shell)}$	4.53	3.88	-	3.85	4.27	3.24
c	$C\alpha$	5.47	4.04	-	3.91	4.61	3.31

Tabla V.5: Resultados de las minimizaciones para el set 3 de estructuras, considerando todas las propiedades conjuntamente.

Set	Modelo	r óptimo (Å) / % dif.			
		a_G	a_T	a_I	a_R
1	Atómico	2.58/3.75 %	2.58/4.18 %	2.58/2.75 %	3.11/5.35 %
	$C\alpha_{(Shell)}$	4.21/3.46 %	4.84/4.25 %	4.53/3.11 %	5.47/4.46 %
	$C\alpha$	-/6.74 %	6.11/4.38 %	5.79/3.28 %	6.74/4.64 %
2	Atómico	-/- %	2.84/4.80 %	2.32/3.76 %	-/- %
	$C\alpha_{(Shell)}$	-/- %	4.84/4.95 %	4.84/3.55 %	-/- %
	$C\alpha$	-/- %	6.11/5.08 %	5.79/3.73 %	-/- %
3	Atómico	-/- %	2.05/3.42 %	2.84/3.58 %	2.32/2.52 %
	$C\alpha_{(Shell)}$	-/- %	4.53/3.85 %	4.84/3.74 %	4.53/3.24 %
	$C\alpha$	-/- %	5.16/3.53 %	6.11/4.11 %	5.47/3.31 %
Global	Atómico	2.58/3.75 %	2.32/4.79 %	2.58/3.73 %	2.84/5.21 %
	$C\alpha_{(Shell)}$	4.21/3.46 %	4.84/4.66 %	4.84/3.54 %	5.47/4.56 %
	$C\alpha$	-/6.74 %	5.79/4.89 %	5.79/3.74 %	6.42/4.72 %

Tabla V.6: Radios óptimos encontrados para cada propiedad considerada, junto al porcentaje de error para esa propiedad, para los tres sets de estructuras.

V.2.2. Proteínas de gran tamaño

Para el caso de proteínas o complejos proteicos de gran tamaño, un factor muy importante a tener en cuenta es el coste computacional que supone calcular las propiedades hidrodinámicas en este tipo de macromoléculas. El programa HYDROPRO genera el modelo de concha con un número de elementos de fricción limitado y menor que 2000, por lo que el tiempo de cálculo y la ocupación de memoria siempre será el mismo, independientemente del tamaño y forma de la macromolécula. Sin embargo, si prescindimos del modelo de concha, y calculamos las propiedades directamente del modelo $C\alpha$, el coste computacional se incrementa hasta hacer impracticable el cálculo, sobre todo en construcciones compuestas por varios miles de aminoácidos. Por el contrario, con el modelo de concha, los cálculos se pueden llevar a cabo en un ordenador de escritorio, sin necesidad de recurrir a máquinas especializadas en cálculo intensivo. Es por ello que el modelado de concha es el método más adecuado para calcular las propiedades hidrodinámicas de proteínas globulares rígidas, independientemente del tamaño y forma de ésta. Para ilustrar esta posibilidad, abordamos el estudio de una serie de proteínas de gran tamaño, con un peso molecular, M , igual o superior a 500000 Da. Para cada una de ellas, calcularemos las propiedades en disolución y se compararán con los datos experimentales disponibles.

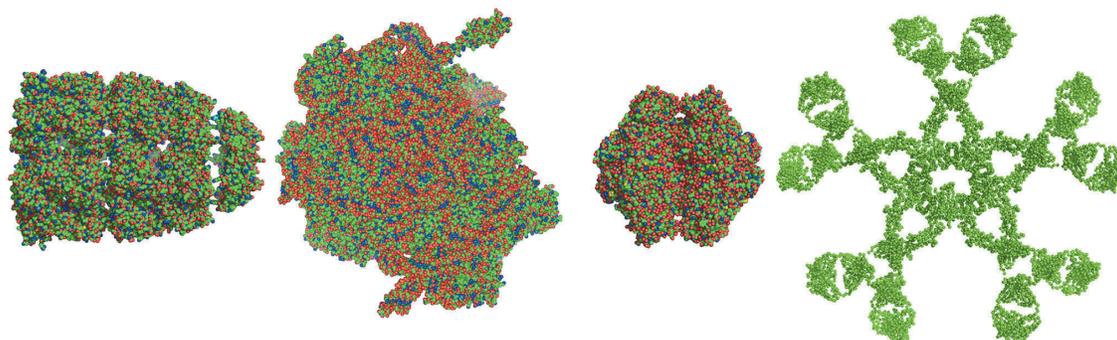


Figura V.5: Estructuras consideradas en este estudio. De izquierda a derecha: GroEL, Ribosoma 70S, Ureasa completa y anticuerpo IgM. De este último solo se dispone de las coordenadas de los átomos $C\alpha$.

Empleamos la misma metodología que en las proteínas de tamaño bajo-medio estudiadas en el caso anterior.

Chaperona GroE

El complejo proteico GroEL-GroES, está compuesto mayoritariamente por la proteína GroEL (7273 aminoácidos, 53480 átomos, con un peso molecular de 802.6 kDa). La estructura del complejo se encuentra disponible en el *Protein Data Bank* bajo el código 2CGT [171], aunque emplearemos únicamente las coordenadas de GroEL. Respecto a las medidas experimentales, conocemos el coeficiente de sedimentación y difusión de GroEL, obtenido de medidas de sedimentación [172] y el radio de giro obtenido de SAXS [173].

Ureasa

Se trata de una enzima dimérica con un peso molecular total de 480000 Da. La estructura atómica de la forma dimérica está disponible (Código PDB: 3LA4 [174]), y se conoce un número elevado de propiedades experimentales, tanto de la forma dimérica como la monomérica [175].

Ribosoma 70S de *E. coli*

La estructura cristalográfica completa del ribosoma 70S de *E. coli*, dividida en dos debido a su gran tamaño, se encuentra depositada en el *Protein Data Bank* bajo los códigos 2AW4 (50S) y 2AVY (30S). Es un complejo ARN-proteína con un peso molecular

de aprox. 2.5×10^6 Da, que se divide a su vez en una subunidad mayor (50S) y otra menor (30S), con pesos moleculares aproximados de 1.7×10^6 Da y 1.0×10^6 Da respectivamente. Realizaremos el cálculo del modelo atómico mediante HYDR0PRO para el ribosoma completo y sus subunidades por separado. Respecto al modelo reducido de elementos centrados en el $C\alpha$, puesto que el ribosoma contiene ARN y proteína, hemos visto conveniente incluir en el modelado hidrodinámico esferas de 3.5 Å centradas en las posiciones de los átomos de fosfato del ácido nucleico, basándonos en un trabajo anterior sobre el modelado hidrodinámico del ADN [176]. Debido a su heterogénea composición, presenta un volumen específico parcial \bar{v} de $0.639 \text{ cm}^3/\text{g}$ [177]. Los datos experimentales son obtenidos de la bibliografía [178].

Anticuerpo IgM

Este anticuerpo, que en su forma natural es un pentámero, forma un complejo con un peso molecular de 840000 Da. La estructura no ha podido ser resuelta completamente, y actualmente solo disponemos de las coordenadas de los átomos $C\alpha$ (Código PDB: 2RCJ), obtenidos de medidas de dispersión de rayos X y modelado molecular [179]. Emplearemos estas coordenadas para el cálculo hidrodinámico. Los datos experimentales son obtenidos de la literatura científica [180].

Resultados

Los datos experimentales disponibles se han resumido en la tabla V.7. Los resultados de los cálculos se muestran en las Tablas V.8, V.9 y V.10. En los tres casos, las propiedades en disolución calculadas están muy próximos a las experimentales. Observamos que la pérdida del detalle atómico no presenta apenas influencia en las propiedades calculadas, debido a las grandes dimensiones de las macromoléculas de este estudio. Respecto al tiempo de cálculo, al emplear el modelo $C\alpha$, que considera cada $C\alpha$ como elemento de fricción, aumenta considerablemente (en algunos casos, más de 20 veces) respecto a los del modelo de concha, con un número fijo de elementos de fricción (2000). Como conclusión, es altamente recomendable emplear el modelado de concha para calcular las propiedades de complejos macromoleculares de gran tamaño, siendo indiferente emplear el modelo atómico o el modelo simplificado de $C\alpha$.

Proteína	M (Da)	\bar{v} (cm ³ /g)	Rg (Å)	Dt $\times 10^7$ (cm ² /s)	$[\eta]$ (cm ³ /g)	s (S)
GroEL	802600 [172]	0.747 [172]	67 [173]	2.59 [172]		22.13 [172]
Ureasa completa	480000 [181]	0.73 [181]		3.46 [181]		18.6 [181]
Ureasa mitad	240000 [182]	0.73 [181]				11.5 [182]
Ig M	950000 [180]	0.722 [183]	121 [184]	1.82 [183]	13.4 [180]	17.5 [183]
Ribosoma 30S	1000000 [185]	0.639 [177]	69 [178]	2.18 [186]	8.1 [178]	31.8 [178]
Ribosoma 50S	1700000 [185]	0.639 [177]	77 [178]	1.90 [186]	5.6 [178]	50.2 [178]
Ribosoma 70S	2500000 [177]	0.639 [177]	125 [178]	1.72 [177]	6.8 [178]	70.5 [178]

Tabla V.7: Datos experimentales de las estructuras de gran tamaño consideradas en este estudio.

Proteína	N_{atomos}	r_{EA} (Å)	t CPU (s)	Rg (Å)	Dt $\times 10^7$ (cm ² /s)	$[\eta]$ (cm ³ /g)	s (S)	τ_h (ns)
GroEL	53466	2.84	360	66.1	2.58	4.54	21.52	626
Ureasa completa	38184	2.84	360	49.17	3.29	3.66	17.96	300
Ureasa mitad	19092	2.84	360	39.9	4.14	3.70	11.3	157
Ribosoma 30S	54884	2.84	360	67.5	2.55	4.2	34.0	720
Ribosoma 50S	91648	2.84	360	71.7	2.22	3.6	52.0	1131
Ribosoma 70S	146532	2.84	360	87.0	1.82	4.1	67.6	1662

Tabla V.8: Resultados del cálculo hidrodinámico del modelo atómico.

Proteína	N_{residuos}	$r_{C\alpha(\text{Shell})}$ (Å)	t CPU (s)	Rg (Å)	Dt $\times 10^7$ (cm ² /s)	$[\eta]$ (cm ³ /g)	s (S)	τ_h (ns)
GroEL	7273	4.84	360	64.4	2.68	4.06	22.3	552
Ureasa completa	4996	4.84	360	48.8	3.31	3.59	18.1	293
Ureasa mitad	2504	4.84	360	40.3	4.12	3.75	11.26	159
Ig M	7514	4.84	360	126.7	1.73	13.7	18.4	2780
Ribosoma 30S	3883	4.84	360	67.4	2.5	4.4	33.4	719
Ribosoma 50S	6336	4.84	360	74.3	2.15	3.9	50.4	1209
Ribosoma 70S	10219	4.84	360	89.3	1.81	4.1	67.2	1731

Tabla V.9: Resultados del cálculo hidrodinámico del modelo $C\alpha(\text{Shell})$.

Proteína	N_{residuos}	$r_{C\alpha}$ (Å)	t CPU (s)	Rg (Å)	Dt $\times 10^7$ (cm ² /s)	$[\eta]$ (cm ³ /g)	s (S)	τ_h (ns)
GroEL	7273	6.11	464	65.0	2.58	4.45	21.54	608
Ureasa completa	4996	6.11	184	47.7	3.27	3.65	17.4	300
Ureasa mitad	2504	6.11	51	39.1	4.01	3.96	10.6	171
Ig M	7514	6.11	430	126.3	1.70	13.76	18.4	2821
Ribosoma 30S	4027	6.11-3.5(P)	110	65.9	2.62	3.8	38.4	661
Ribosoma 50S	6401	6.11-3.5(P)	320	77.6	2.21	3.6	50.8	1021
Ribosoma 70S	10428	6.11-3.5(P)	1720	91.0	1.66	4.6	74.9	1574

Tabla V.10: Resultados del cálculo hidrodinámico del modelo $C\alpha$.

V.3. Estudio conformacional e hidrodinámico de proteínas desnaturalizadas

Para poder realizar un estudio preciso de la conformación y dinámica global e interna en disolución de macromoléculas, es preciso trabajar en escalas de tiempo grandes, del orden de microsegundos. Como es bien conocido, la técnica de simulación de dinámica molecular atómica (DM) no es practicable para estudiar la evolución de estructuras macromoleculares a estas escalas de tiempo, por lo que debemos emplear modelos simplificados (o reducidos) y procedimientos de simulación alternativos, como son la dinámica browniana (DB) y Monte Carlo (MC). Por lo tanto, desarrollar metodologías que permitan estudiar la estructura global y el comportamiento mediante simulación computacional, basada en modelado hidrodinámico y centrada en tratar de reproducir sus propiedades en disolución, puede suponer un avance importante y un apoyo de gran valor para las técnicas experimentales. A continuación, describiremos una metodología computacional adecuada para el estudio de proteínas, desde el punto de vista de la conformación e hidrodinámica.

En primer lugar, abordamos el estudio con proteínas desplegadas en presencia de agentes desnaturalizantes, alta temperatura o en medio ácido, siempre en presencia de agentes reductores, que rompen los puentes disulfuro intramoleculares, lo que nos permite parametrizar el modelo flexible. Posteriormente, realizaremos simulaciones de proteínas intrínsecamente desordenadas, concretamente, aquellas compuestas por dominios globulares rígidos unidos por segmentos flexibles, empleando las parametrizaciones halladas para modelos reducidos de proteínas globulares y proteínas desnaturalizadas.

V.3.1. Modelo de simulación de proteína desnaturalizada

En nuestro modelo, cada aminoácido es representado como una única esfera centrada en el átomo $C\alpha$, que engloba la cadena lateral; así, optamos por el modelo más básico, sin elementos adicionales para la cadena lateral. Básicamente, se trata de una cadena compuesta por N elementos y $N - 1$ enlaces, con una distancia de enlace centrada en 3.8 Å. Para definir las interacciones angulares y torsionales, se realizó un análisis estadístico en estructuras tridimensionales de proteínas, pertenecientes al *Protein Data Bank*.

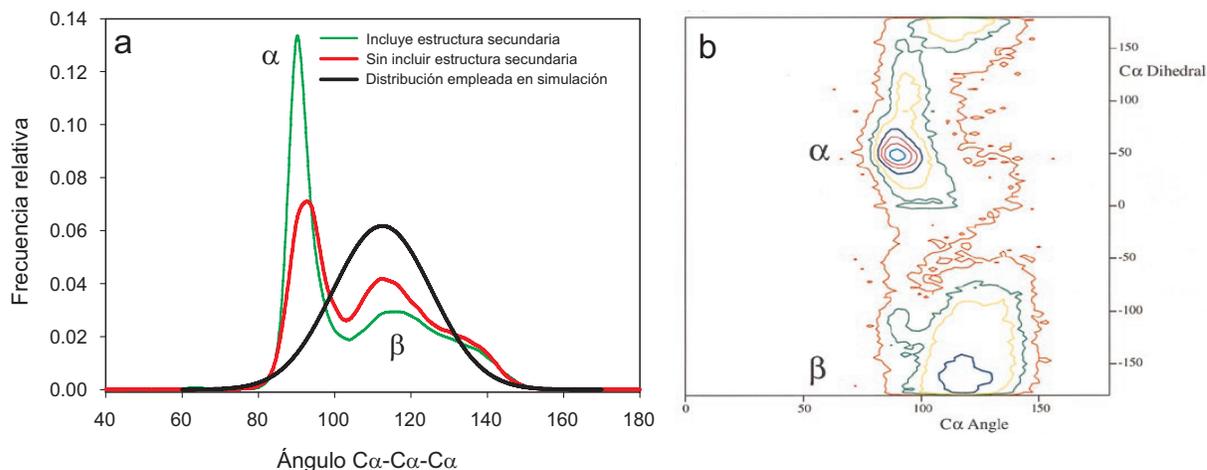


Figura V.6: Distribuciones obtenidas del análisis estadístico de la selección de estructuras depositadas en el *Protein Data Bank* de estructuras completas (línea verde), regiones sin estructura asignada (línea roja) y las observadas en las simulaciones (línea negra), para los ángulos de enlace. (d) Mapa de curvas de nivel obtenida de Kleywegt *et al* [187].

Análisis estadístico de estructuras de proteínas

La base de datos seleccionada corresponde a la seleccionada por el Prof. Gert Vriend [188], denominada *PDB-Select*, alojada en <http://swift.cmbi.ru.nl/gv/select>. En la fecha de publicación de esta Tesis, está compuesta por una selección de 2710 estructuras no redundantes, que comparten un mínimo de homología, y cuya precisión es alta (R-factor $< 0.19 \text{ \AA}$ y resolución $< 1.9 \text{ \AA}$). Se analizaron por un lado el conjunto completo de estructuras, y por otro lado, un subconjunto que corresponde a las regiones desestructuradas, como bucles o tallos terminales. Se almacenaron los ángulos entre tres $C\alpha$ consecutivos y los ángulos diedros, entre cuatro $C\alpha$ consecutivos.

Como resultados del estudio, observamos que la inmensa mayoría de los ángulos analizados entre tres $C\alpha$ consecutivos varía aproximadamente entre 75° y 150° en ambos casos, con una distribución característica (ver Figura V.6a). Si incluimos en el análisis las estructuras secundarias, se observa un pico prominente alrededor de los 90° que corresponde al ángulo característico de la hélice- α , la más frecuente, mientras que alrededor de 120° , se observa otro pico correspondiente al ángulo característico de la hoja- β , solapado con disposiciones extendidas en el rango $120\text{-}140^\circ$ [66]. Al excluir los ángulos de las estructuras secundarias, se observa una distribución muy parecida a la anterior, con la diferencia de que dichos picos mayoritarios aparecen casi a la misma frecuencia. Respecto a los ángulos

diedros, vemos una distribución muy homogénea, donde los ángulos diedros prácticamente pueden tomar cualquier valor. Estos resultados son confirmados en un trabajo de Kleywegt *et al* [187], donde se realizó un análisis estadístico similar de los ángulos de enlace y diedros entre átomos $C\alpha$ en estructuras completas, aunque en ese estudio están presentes las estructuras secundarias (ver Figura V.6b).

Por consiguiente, decidimos emplear un potencial armónico simple (ver Figura V.6a, línea negra) con el fin de representar la distribución angular observada experimentalmente mediante una distribución gaussiana convencional, que cubra todo el rango de ángulos permitidos. Respecto a los ángulos de torsión, decidimos no incluir ninguna interacción para los ángulos diedros.

V.3.2. Propiedades conformacionales de proteínas desnaturalizadas

A continuación, se procede a la parametrización conformacional del modelo de proteína desnaturalizada, cuyo único parámetro ajustable es el radio de esferas duras para representar el volumen excluido.

Mediante Monte Carlo se generaron cadenas de N elementos, con un procedimiento muy similar a la construcción de cadenas vermiformes, descrito en la subsección IV.2.3 de la Tesis. Básicamente, las cadenas se generan a partir de ángulos obtenidos con un generador de aleatorios de distribución gaussiana. Para cubrir el rango 75-150° antes citado, en la generación de aleatorios se toma como valor medio $(75 + 150)/2 = 112.5^\circ$, y una desviación estándar tal que el triple de ella cubra, con un intervalo de confianza del 97%, el semi-intervalo $(150 - 75/2) = 37.5^\circ$, siendo dicha desviación estándar $37.5/3 = 12.5^\circ$. Así, la densidad de probabilidad del ángulo, en grados es:

$$p(\alpha) = K \exp\left[-\left(\frac{\alpha - 112.5}{12.5}\right)^2\right] \quad (\text{V.2})$$

Además, el programa incluye la posibilidad de rechazar cadenas que no cumplan la restricciones de volumen excluido. Representamos este efecto de la manera más sencilla, mediante un potencial de esferas duras, que prohíbe que la distancia entre dos esferas no consecutivas sea inferior al doble de un radio de esfera dura, r_{HS} .

Para parametrizar el modelo de proteína flexible, nos basaremos en datos experimentales de radios de giro de proteínas desnaturalizadas. En un trabajo de Kohn *et al* [189] aparece publicada una sistemática recopilación de datos de R_g , desde $N = 8$ hasta

$N = 549$ residuos, obtenidos de SAXS de un gran número de proteínas desplegadas con agentes desnaturalizantes, como urea o hidrocloreuro de guanidinio (GuHCl), en presencia de agentes reductores que rompen y evitan la formación de puentes disulfuro, además de encontrarse libres de grupos prostéticos. La parametrización conformacional consiste en variar el parámetro que define el volumen excluido, es decir, el radio de esfera dura, al que llamaremos en lo sucesivo r_{HS} , y comparar los datos obtenidos, al promediar sobre un conjunto de 10000 cadenas, mediante cálculo Monte Carlo con los datos experimentales. Para ello, se minimiza el error relativo cuadrático medio, obtenido mediante la función Δ ,

$$\Delta = \sqrt{\frac{1}{N_{prots}} \sum_{N_{prots}} \left[\frac{R_g(cal) - R_g(exp)}{R_g(exp)} \right]^2} \quad (V.3)$$

siendo Δ el error relativo promediado, y 100Δ el porcentaje de error relativo hallado. Debido a que esta función puede sobreestimar las diferencias para bajo N , empleamos otra función, que denominamos δ , para hallar el error absoluto del ajuste:

$$\delta = \sqrt{\frac{1}{N_{prots}} \sum_{N_{prots}} \left[R_g(calc) - R_g(exp) \right]^2} \quad (V.4)$$

Sin embargo, debido a que empleamos el método Monte Carlo clásico basado en la cadena de Schellman para generar las cadenas, a medida que aumentamos la longitud de la cadena se incrementa exponencialmente el tiempo de CPU, haciéndose impracticable por encima de $N = 200$. Para poder obtener las propiedades del modelo por encima de 200 residuos, es necesario recurrir a otras estrategias tales como MonteCarlo-Metrópolis, métodos dinámicos u otros.

El resumen de los resultados se recoge en la Tabla V.11 y Figura V.7. De los diversos radios r_{HS} que se probaron (1.9, 2.1, 2.25, 2.4 y 2.5 Å), el que mejor ajusta a los datos experimentales es el valor de $r_{HS} = 2.25$ Å. Es destacable que el ajuste sigue, en todo el rango, la tendencia de los valores experimentales ajustándolos correctamente: el error típico relativo es menor del 5% y el absoluto de 1.4 Å, que son del orden de los típicos errores experimentales y de la fluctuación entre unos y otros valores del grupo de datos. Además, en la Figura V.7, representamos también la ecuación de ajuste a los datos experimentales propuesta por el mismo autor del trabajo de recopilación de los radios de giro. Observamos que la ecuación ($R_g = 1.927N^{0.589}$) prácticamente se superpone sobre la curva de datos calculados con $r_{HS} = 2.25$ Å, pese a que esta ecuación de ajuste se realizó

V.3. ESTUDIO CONFORMACIONAL E HIDRODINÁMICO DE PROTEÍNAS DESNATURALIZADAS

Proteína	N	$\langle R_g \rangle$	Desnaturalizante	$r_{HS} = 2.1 \text{ \AA}$		$r_{HS} = 2.25 \text{ \AA}$		$r_{HS} = 2.4 \text{ \AA}$	
				$\langle R_g \rangle$	dif.	$\langle R_g \rangle$	dif.	$\langle R_g \rangle$	dif.
AK-16	16	9.8	4M urea	9.38	-0.42	9.48	-0.32	9.61	-0.19
AK-27	27	12.8	6M GuHCl	13.16	0.36	13.33	0.53	13.53	0.73
AK-32	32	14.5	6M GuHCl	14.68	0.18	14.90	0.40	15.08	0.58
AK-37	37	16.9	6M GuHCl	16.15	-0.75	16.34	-0.56	16.59	-0.31
N-ter citocromo c	39	18.4	4M urea	16.68	-1.72	16.89	-1.51	17.17	-1.23
Proteína G	52	23.0	2.3M GuHCl	20.00	-3.00	20.35	-2.65	20.63	-2.37
drK SH3	59	21.9	2M GuHCl	21.76	-0.14	22.04	0.14	22.44	0.54
GCN4-p2'	66	24.1	4.26M GuHCl	23.32	-0.78	23.68	-0.42	24.04	-0.06
Ubiquitina	76	25.2	4.96M GuHCl	25.41	0.21	25.82	0.62	26.28	1.08
Fyn SH3	78	25.7	6M GuHCl	25.81	0.11	26.22	0.52	26.76	1.06
Proteína L	79	26.0	45M GuHCl	25.96	-0.04	26.49	0.49	26.94	0.94
mAcP	98	30.4	6.58M urea	29.90	-0.50	30.22	-0.18	30.87	0.47
ctACP	98	30.5	8M urea	29.90	-0.60	30.22	-0.28	30.87	0.37
p13K SH3	103	30.9	2.67M GuHCl	30.74	-0.16	31.35	0.45	31.93	1.03
p13K SH2	112	29.6	3M GuHCl	32.46	2.86	32.94	3.34	33.54	3.94
RNaseA	124	33.2	3.2M GuHCl	34.52	1.32	35.17	1.97	35.83	2.63
Lisozima	129	35.8	4M GuHCl	35.28	-0.52	35.85	0.05	36.57	0.77
CheY	129	38.0	5-7 M urea	35.28	-2.72	35.85	-2.15	36.57	-1.43
Snasa	149	37.2	6M GuHCl	38.63	1.43	39.24	2.04	40.04	2.84
Apo-mioglobina	154	40.0	6M GuHCl	39.21	-0.79	40.13	0.13	41.01	1.01
DHFR	167	44.0	8M urea	41.41	-2.59	42.08	-1.92	42.90	-1.10
Δ (% error relativo)					5.07		4.94		5.37
δ (error absoluto / \AA)					1.40		1.36		1.52

Tabla V.11: Selección de resultados de la parametrización del radio de esfera dura r_{HS} . Se expresa en la Tabla resultados de R_g (en \AA) para tres valores diferentes de r_{HS} junto con la diferencia (dif., en \AA) con respecto al valor experimental de un conjunto de proteínas desnaturalizadas. Al final de la tabla se muestra el porcentaje de error relativo y absoluto hallado.

para N de hasta 549 residuos, confirmando el excelente ajuste a datos experimentales de proteínas desnaturalizadas logrado con nuestro modelo de proteína desordenada.

Como hemos visto, el modelo de proteína flexible está basado en datos de fragmentos de proteínas globulares de estructura conocida en condiciones no-desnaturalizantes, cuya estructura se pierde en presencia de agentes desnaturalizantes y reductores. Sin embargo, como se comenta en la introducción de este capítulo, existe otro tipo de proteínas que no presentan ningún tipo de estructura definida en condiciones fisiológicas, a las que denominamos proteínas intrínsecamente desordenadas (PID), cuya naturaleza en un principio podría considerarse más próxima a las regiones flexibles de sistemas mixtos de proteínas flexibles, cuya simulación es el objetivo principal de este capítulo (serán tratadas extensamente más tarde). Mientras que algunas PID poseen regiones globulares rígidas bien definidas, otras se supone que están predominantemente desordenadas en toda su extensión. Cuando estas últimas son introducidas en el análisis se observa una gran dispersión (ver Figura V.7, puntos verdes) en comparación con los datos de proteínas desnaturalizadas. Empleando la función Δ , analizando en conjunto los R_g de proteínas desnaturalizadas e intrínsecamente desordenadas, obtenemos un error global del 10.4%, siendo del 14%

cuando analizamos solo las intrínsecamente desordenadas. Los errores relativos (función δ) son de 3.2 Å y 4.5 Å respectivamente. Aparentemente, se podría mejorar el ajuste empleando un r_{HS} menor. Con $r_{HS} = 2.10$ Å, obtenemos un 9.8% ($\delta = 3.0$ Å) de error global para desnaturalizadas e intrínsecamente desordenadas, y 13.3% ($\delta = 4.2$ Å) cuando se analizan solo las intrínsecamente desordenadas. Sin embargo, apenas se mejora el ajuste debido a la alta dispersión de los datos.

Para encontrar una posible explicación a la gran fluctuación observada en los radios de giro de estas PID, debemos referirnos a la naturaleza física de este tipo de proteínas, que suelen presentar un alto contenido en residuos cargados, por lo que presentan una carga elevada a pH fisiológico. Por ejemplo, la proteína α -protomiosina presenta un 49% de residuos cargados positivamente (ácidos Asp y Glu) y otras proteínas, como las sinucleínas (α, β y γ) presentan una carga neta anormalmente elevada a pH fisiológico, por lo que estas proteínas se encuentran expandidas respecto a una proteína químicamente desnaturalizada con el mismo número de residuos [189]. Es destacable el notable descenso que se produce en el radio de giro de α -protomiosina y α -sinucleína cuando pasamos de un pH fisiológico a un pH ácido (ver Tabla V.12), lo que confirma el efecto de expansión debido a la repulsión entre aminoácidos cargados, que desaparece al protonarse estos. Tampoco podemos descartar con absoluta certeza que no exista algún tipo de estructura secundaria residual en este tipo de proteínas, por ejemplo en las construcciones de la proteína Tau [190]. Por otro lado, es bien conocido que la presencia de agentes desnaturalizantes a alta concentración y/o reductores imposibilitan la formación de estructuras secundarias, además de apantallar las posibles cargas que puedan presentar las cadenas proteicas, evitando las posibles interacciones atractivas o repulsivas entre los residuos cargados.

Hemos considerado conveniente mostrar datos de radios de giro de proteínas PID recopilados de la bibliografía [190], representados junto a las proteínas desnaturalizadas químicamente y los datos calculados en la figura V.7, con el objeto de mostrar como se desvía notablemente de los de las desnaturalizadas. No obstante, adoptamos como parametrización la resultante de los datos de estas últimas solamente, pues presentan total desorden y flexibilidad. Esto es coherente con el objetivo de esta parametrización, que es obtener un modelo aproximado para la simulación de fragmentos flexibles en sistemas mixtos, donde coexisten en la misma molécula regiones rígidas y regiones flexibles, siendo estas últimas de un tamaño relativamente corto, frecuentemente de menos de 200 aminoácidos.

V.3. ESTUDIO CONFORMACIONAL E HIDRODINÁMICO DE PROTEÍNAS DESNATURALIZADAS

Proteína	N	$\langle R_g \rangle$ (Å)
Dominio PIR	75	26.5
p53 (1-93)	93	28.7
Tau K19	99	35
Protimosina α pH 7.5	109	37.8
Protimosina α pH 2.5	109	27.6
Neuroigina	118	31.5
Tau K18	130	38
Tau K18	130	35
N-tail nucleoproteina	139	27.2
α -Sinucleina pH 7.5	140	40
α -Sinucleina pH 3.0	140	30
Tau K17	143	36
HrpO	147	35
Calpastatina	148	35.4
Tau K10	167	40
Tau K27	171	37
Tau K16	174	39
Tau K25	185	41

Tabla V.12: Recopilación de radios de giro de proteínas intrínsecamente desordenadas (PID), obtenidos de Bernadó *et al* [190]. N indica el número de residuos (elementos) de la cadena.

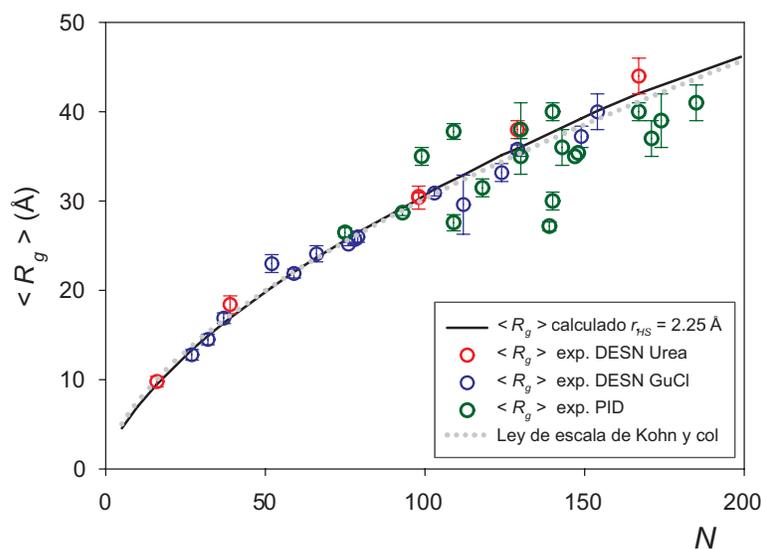


Figura V.7: Representación de los datos experimentales de proteínas desnaturalizadas (DESN) en urea (puntos rojos) y GuCl (puntos azules) junto a la ley de escala propuesta por Kohn *et al* (línea discontinua) y los radios de giro calculados del modelo (línea continua). Se representan además las proteínas intrínsecamente desordenadas (PID) (puntos verdes).

V.3.3. Propiedades hidrodinámicas de proteínas desnaturalizadas

A continuación procedemos a evaluar el modelo de proteína desordenada en lo que a propiedades hidrodinámicas se refiere. Disponemos de una amplia colección de propiedades de transporte experimentales, desde un número de aminoácidos $N = 16$ hasta $N = 200$, obtenidos de la bibliografía, para proteínas desnaturalizadas químicamente, bien en presencia de agentes desnaturalizantes o en medio ácido. Por un lado, tenemos datos provenientes de coeficientes de sedimentación s y viscosidad intrínseca $[\eta]$ [191, 192, 193, 194], coeficientes de difusión traslacional D_t [195, 196, 197, 198, 199] y por último, radios hidrodinámicos procedentes de cromatografía de exclusión por tamaño (SEC) [200], aunque en este último caso, no se trata de determinaciones directas, sino estimadas a partir del volumen de elución, cuya calibración se realizó con proteínas globulares estructuradas. Los valores están presentados como radios equivalentes en las Tablas V.13 para a_T y V.14 para a_I . También se han incluido para su análisis datos experimentales de proteínas intrínsecamente desordenadas, obtenidos de una recopilación realizada por Uversky *et al* [201], en su mayoría radios hidrodinámicos procedentes de cromatografía de exclusión. Estos valores se muestran en las Tablas V.13 para a_T y V.15 para a_I .

Hay que notar que los valores procedentes de las magnitudes traslacionales, s y D_t , suministran a_T , mientras que los procedentes de $[\eta]$ suministran a_I , usando las ecuaciones descritas en la subsección II.1.3, donde $a_T = KT/6\pi\eta_0 D_t$ y $a_I = \sqrt[3]{[\eta]3M/10\pi N_A}$.

La metodología de Monte Carlo es equivalente a la realizada para la parametrización conformacional, con el añadido del cálculo hidrodinámico, realizando un promedio sobre 10000 cadenas generadas. Para ello, debemos asignar un radio hidrodinámico a las esferas que componen el modelo. Se emplea el radio anteriormente hallado para proteínas globulares rígidas modeladas a nivel de residuo (sección V.2), con átomos centrados en las posiciones $C\alpha$, donde $r = 6.11$ Å. Nótese que este dato se usa para las proteínas flexibles no como un parámetro ajustable, sino como un valor adoptado coherentemente del estudio previo de las proteínas rígidas.

Por conveniencia, las propiedades experimentales y calculadas mediante Monte Carlo se comparan en forma de radios equivalentes. Los datos experimentales, en el intervalo $N = 10 - 200$, son transformados a radios equivalentes y son comparados frente a los de la simulación, empleando para ello el método basado en la función Δ (ec. V.3), anteriormente utilizada en la parametrización de R_g .

V.3. ESTUDIO CONFORMACIONAL E HIDRODINÁMICO DE PROTEÍNAS DESNATURALIZADAS

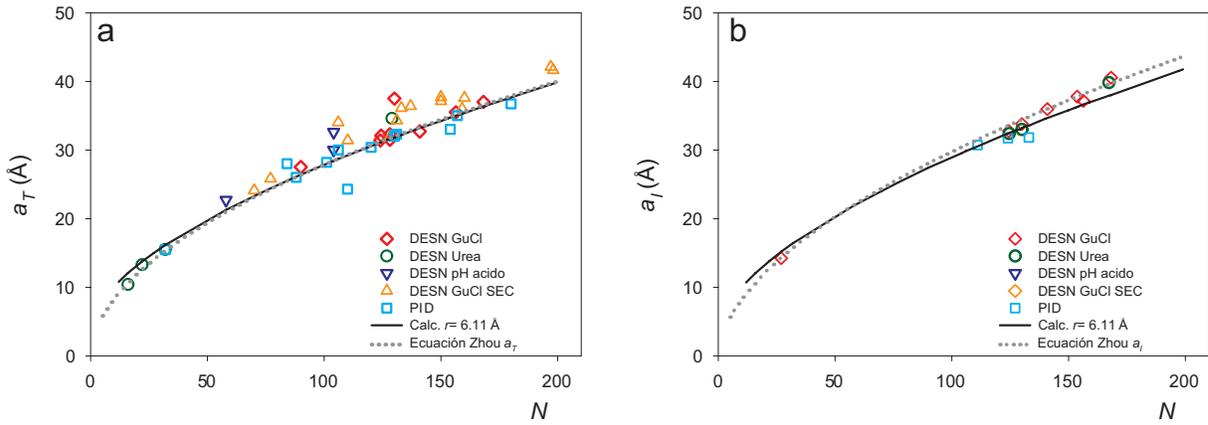


Figura V.8: Resultados de la parametrización hidrodinámica del modelo. (a) Representación de los datos experimentales de proteínas desnaturalizadas (DESN) e intrínsecamente desordenadas (PID) junto a la ley de escala propuesta por Zhou *et al* [195] para a_T (línea discontinua) y los radios equivalentes a_T calculados del modelo (línea continua). (b) Representación de los datos experimentales de proteínas desnaturalizadas (DESN) e intrínsecamente desordenadas (PID), junto a la ley de escala propuesta por Zhou *et al* [195] para a_I (línea discontinua) y los radios equivalentes a_I calculados del modelo (línea continua)

En la Figura V.8 podemos observar los resultados del análisis. Desde el punto de vista cuantitativo, empleando la función Δ , hallamos un error del 7.5% para a_T ($\delta = 2.4$ Å) y 3.8% para a_I ($\delta = 1.27$ Å). Al incluir las intrínsecamente desordenadas en el análisis, obtenemos un error del 7.3% para a_T ($\delta = 2.26$ Å) y 3.7% para a_I ($\delta = 1.23$ Å). Y por último, si realizamos el análisis solo con las intrínsecamente desordenadas, obtenemos un error del 6.8% para a_T ($\delta = 0.07$ Å) y 3.5% para a_I ($\delta = 1.1$ Å). Además se incluyen las leyes de escala empíricas para proteínas desnaturalizadas halladas por Zhou *et al* [195] ($a_T = 3.53N^{0.449}$) para las propiedades traslacionales y para la viscosidad intrínseca ($[\eta] = 0.703N^{0.666}$). Se observa una buena superposición de ambas curvas (ley de escala y calculada por Monte Carlo), especialmente en el caso de a_T , en el rango de $N < 200$ residuos.

N	Ref.	Proteína	Disolvente	Técnica	a_T (Å)
130	[197]	Lisozima	GuHCl 6M	DLS	37.50
128	[198]	α -Lactoalbúmina (bov.)	GuHCl 6M	SAXS	31.50
128	[199]	α -Lactoalbúmina (bov.)	GuHCl 6M	AUC	32.30
124	[191]	Ribonucleasa A (1)	GuHCl 6M	AUC	32.08
140	[191]	Hemoglobina	GuHCl 6M	AUC	32.70
156	[191]	Mioglobina	GuHCl 6M	AUC	35.50
168	[191]	β -Lactoglobulina	GuHCl 6M	AUC	37.00
16	[196]	Lisozima 49-64	Urea 8M	PFG NMR	10.40
22	[196]	D2FPB 17-37	Urea 8M	PFG NMR	13.30
32	[196]	D2FPB 2-38	Urea 8M	PFG NMR	15.50
129	[196]	Hen Lisozima	Urea 8M	PFG NMR	34.60
90	[196]	SH3 PI3K	GuHCl 6M	PFG NMR	27.50
104	[196]	Citocromo c caballo	pH ácido	PFG NMR	32.60
58	[195]	BPTI	pH ácido	PFG NMR	22.70
104	[195]	Apo-citocromo c caballo	pH ácido	DLS	30.00
124	[195]	Ribonucleasa A (2)	GuHCl 6M	DLS	31.40
154	[200]	Apo-mioglobina	GuHCl 6M	DLS	43.00
70	[200]	Albebetina	GuHCl 6M	SEC	24.10
77	[200]	Ubiquitina	GuHCl 6M	SEC	25.80
106	[200]	Apo-citocromo c	GuHCl 6M	SEC	34.00
110	[200]	Protimosina	GuHCl 6M	SEC	31.40
128	[200]	α -Lactoalbúmina	GuHCl 6M	SEC	31.80
131	[200]	Sinucleina	GuHCl 6M	SEC	34.30
133	[200]	Intestinal fatty acid-binding 1	GuHCl 6M	SEC	36.10
137	[200]	Intestinal fatty acid-binding 2	GuHCl 6M	SEC	36.40
150	[200]	Apoflavodoxina	GuHCl 6M	SEC	37.70
150	[200]	p16	GuHCl 6M	SEC	37.10
159	[200]	Nucleasa de estafilococos	GuHCl 6M	SEC	36.00
160	[200]	Dihidrofolato reductasa	GuHCl 6M	SEC	37.60
197	[200]	Adenilato ciclasa	GuHCl 6M	SEC	42.10
198	[200]	Lactógeno de placenta (bov.)	GuHCl 6M	SEC	41.60

Tabla V.13: Recopilación de datos experimentales a_T de proteínas desnaturalizadas.

N	Ref.	Proteína	Disolvente	Técnica	a_I (Å)
130	[192]	Lisozima	GuHCl 6M	Viscosimetría	33.78
130	[192]	Sinucleina	Urea	Viscosimetría	32.97
167	[193]	β -Lactoglobulina	Urea	Viscosimetría	39.80
124	[193]	Ribonucleasa A (1)	Urea	Viscosimetría	32.36
153	[194]	Mioglobina	GuHCl 6M	Viscosimetría	37.77
27	[191]	Insulina	GuHCl 6M	Viscosimetría	14.22
124	[191]	Ribonucleasa A (2)	GuHCl 6M	Viscosimetría	32.75
140	[191]	Hemoglobina	GuHCl 6M	Viscosimetría	35.95
156	[191]	Mioglobina	GuHCl 6M	Viscosimetría	37.15
168	[191]	β -Lactoglobulina	GuHCl 6M	Viscosimetría	40.53

Tabla V.14: Recopilación de datos experimentales a_I de proteínas desnaturalizadas.

V.3. ESTUDIO CONFORMACIONAL E HIDRODINÁMICO DE PROTEÍNAS DESNATURALIZADAS

N	Autor	Proteína	Técnica	a_T (Å)
32	[201]	α -Fetoproteína, fragmento	SEC	15.5
84	[201]	Vmw65 C-terminal	SEC	28.0
88	[201]	PDE c	SEC	26.0
101	[201]	Em	SEC	28.2
106	[201]	Apo-citocromo c	DLS	30.0
110	[201]	Protimosina α	SEC	24.3
120	[201]	γ -Sinucleína	SEC	30.4
130	[201]	β -Sinucleína	SEC	32.0
131	[201]	α -Sinucleína	SEC	32.3
154	[201]	Estamina	SEC	33.0
157	[201]	CFos-AD	SEC	35.0
180	[201]	Histona de timo	SEC	36.7

N	Autor	Proteína	Técnica	a_I (Å)
111	[201]	Dom. unión fibronectina B	Viscosim.	30.7
124	[201]	Dom. unión fibronectina A	Viscosim.	31.7
133	[201]	Dom. unión fibronectina D	Viscosim.	31.8

Tabla V.15: Recopilación de datos experimentales a_T y a_I de proteínas intrínsecamente desordenadas.

Relación de ratios XY

A continuación, mediante las ecuaciones de ajuste empíricas para los radios a_X (ecuaciones de Kohn *et al* [189] para a_G y Zhou *et al* [195] para a_T y a_I) evaluamos las relaciones XY , concretamente GT , GI e IT , que se presentan en la Figura V.9. Es de notar en primer lugar que los valores de IT son próximos a la unidad, como para cualquier otro tipo de conformación macromolecular, lo cual confirma la coherencia de los valores de D_t o s con los de $[\eta]$. Lo más notable es que los radios GT y GI son apreciablemente inferiores a los valores típicos de polímeros flexibles ($GT = 1.65 - 1.87$, $GI = 1.53 - 1.69$, correspondiendo los límites a condiciones *theta* y buen disolvente, respectivamente). Algo similar ha sido notado recientemente por Berry [202], que utilizó otras combinaciones universales y adimensionales, de R_g , D_t y $[\eta]$, diferentes en forma pero del mismo tipo que GT y GI , encontrando valores diferentes a los de las cadenas lineales, y atribuyéndolo a efectos de no-linealidad, tales como ciclación transitoria por formación temporal de contactos intermoleculares o formación de otros tipos de estructura local. Nuestros resultados indican que tales efectos no son necesarios, ya que la propia cadena lineal los produce. Nuestra interpretación se dirige a la rigidez local de la cadena polipeptídica, y que las cadenas con unos pocos cientos de aminoácidos no son suficientemente largas (en términos de un modelo vermiforme, L pequeña, P grande y por consiguiente, pequeño L/P) para que se apliquen los valores límite.

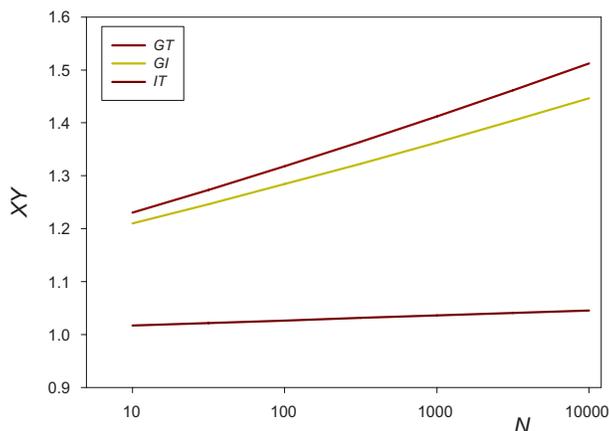


Figura V.9: Representación de las relaciones de radios XY de las propiedades experimentales, obtenidas mediante regresión lineal de las leyes de escala para los radios equivalentes.

V.4. Simulación de proteínas parcialmente desordenadas

Como comentamos anteriormente, uno de los objetivos de este capítulo es desarrollar una metodología computacional aplicable al estudio conformacional y dinámico de proteínas parcialmente desordenadas. Como bien es sabido, este tipo de macromoléculas consta de regiones que presentan estructuras bien definidas (plegadas) y regiones desordenadas, que aparentan un comportamiento aleatorio unido a una gran flexibilidad. Para simular las zonas flexibles nos basaremos en el modelo y la parametrización anterior para proteínas desnaturalizadas con agentes químicos. Para las zonas plegadas, emplearemos el modelo tipo $G\bar{o}$ [65]. A continuación se muestra un breve resumen de los modelos y potenciales empleados.

V.4.1. Simulación de fragmentos flexibles

Los potenciales de enlace, angular, diedro y entre elementos no-enlazados son los siguientes:

- *Potencial de enlace.* La distancia entre $C\alpha$ debe mantenerse fija a una distancia de 3.8 Å. Emplearemos un potencial armónico simple (ec. V.5) con una constante de energía alta $K_i = 100.0 \text{ Kcal/mol/\AA}^2$ y una distancia de equilibrio de 3.8 Å.

$$V(b_i) = K_i(r_i - r_0)^2 \quad i = 1, \dots, N - 1 \quad (\text{V.5})$$

- *Potencial de ángulo de enlace.* Para poder reproducir la distribución gaussiana anteriormente descrita, emplearemos un potencial angular armónico simple (ec. V.6) con una constante de energía $Q_i = 1.77 \text{ Kcal/mol/rad}^2$ y un ángulo de equilibrio interno de 112.5° . EL valor de Q_i se corresponde con la desviación estándar de 12.5° indicada en la sección V.3.2.

$$V(\alpha_i) = Q_i(\alpha_i - \alpha_{i0})^2 \quad i = 1, \dots, N - 2 \quad (\text{V.6})$$

- *Potencial de ángulo diedro.* Debido a que se observa experimentalmente que el ángulo diedro presenta una amplia variación, y prácticamente puede tomar cualquier valor, no definimos ningún potencial.
- *Potencial de elementos no-enlazados.* Para incluir el efecto del volumen excluido en los cálculos Monte Carlo, el radio de esfera dura r_{HS} hallado en la parametrización será incluido mediante un potencial de esfera dura discontinuo, con una distancia límite de $2 \times r_{HS} = 4.5 \text{ \AA}$, como se indicó en la sección V.3.2. Este potencial se emplea en la parametrización mediante Monte Carlo, y en las simulaciones MC-Metrópolis:

$$V_{r_{ij}} = \begin{cases} K_{HS}, & \text{si } r_{ij} \leq r_{cutoff} \\ 0, & \text{si } r_{ij} > r_{cutoff} \end{cases} \quad (\text{V.7})$$

Pero debido a que en dinámica browniana no es posible emplear potenciales discontinuos, utilizaremos en su lugar para elementos no enlazados, un potencial continuo repulsivo, perteneciente al modelo tipo Gō:

$$V_{ne}^{no-nativo}(i, j) = \epsilon_2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12}, \quad \text{si } r_{ij} \text{ es un contacto no-nativo} \quad (\text{V.8})$$

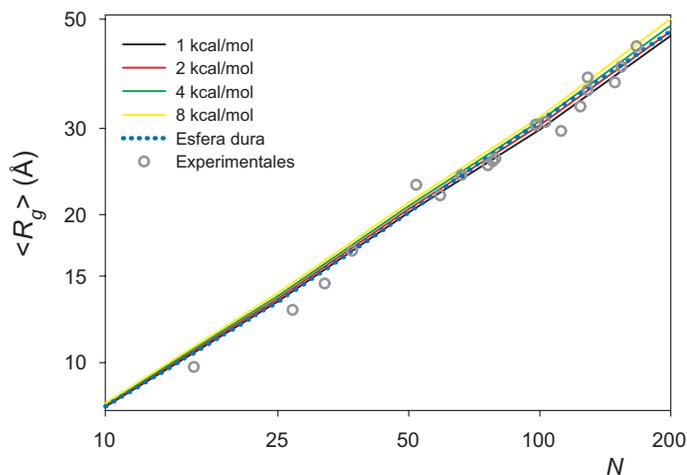


Figura V.10: Representación del radio de giro obtenido mediante el uso de potenciales continuos con diferentes constantes de energía ϵ_2 en dinámica browniana según la ecuación V.8 (líneas continuas), frente al obtenido anteriormente con el potencial de esfera dura en Monte Carlo-CARP según la ecuación V.7 (línea discontinua). Los puntos grises corresponden a los puntos experimentales.

Para parametrizar este potencial, se realizaron una serie de simulaciones de $20 \mu\text{s}$ para $N=10, 25, 100, 150$ y 200 , variando el término energético ϵ_2 entre 1-8 kcal/mol, del cual depende principalmente la fuerza de repulsión entre las esferas. Para el parámetro σ_{ij} , se asignó un valor de 4 \AA , el mismo que el modelo tipo Gō original. Los resultados se muestran en la Figura V.10. Se obtuvieron valores promedio muy similares a los experimentales en los radios de giro calculados para diversos N para todos los parámetros probados, especialmente en el rango de $\epsilon_2 = 1-4 \text{ Kcal/mol}$. Sin embargo, resultó ser el parámetro $\epsilon_2 = 2.0 \text{ Kcal/mol}$ el que mejor ajusta al potencial de esfera dura para todo el rango de N .

V.4.2. Simulación de fragmentos estructurados

La estructura que adopta una proteína plegada es producto de las interacciones específicas entre los aminoácidos que la componen. Sin entrar en detalles, detectar estas interacciones en una estructura estática (cristalografía R-X) o quasi-estática (confórmeros de NMR) no es una tarea trivial. Debemos tener en cuenta la proximidad y disposición geométrica de los pares de aminoácidos en contacto, la superficie de contacto y la naturaleza del mismo. Por ello, es necesario emplear programas específicos que determinen el

número y la naturaleza de las interacciones entre los aminoácidos que mantienen plegada la estructura, de la manera más realista posible.

Para analizar los tipos y cantidad de contactos generados en la estructura atómica emplearemos el programa CSU (Contacts of Structural Units) [203], el cual evalúa los contactos interatómicos en biomoléculas y sus complejos.

Metodología CSU

En primer lugar, el programa analiza las distancias entre los centros de masas de las cadenas laterales de los aminoácidos. Esta distancia debe ser menor a un determinado valor, que denominamos r_{SC} , que oscila entre 4.5-6.5 Å según los autores que emplean este modelo [65, 204]. Si para un determinado par de aminoácidos, la distancia entre sus cadenas laterales es menor que r_{SC} estos se consideran en contacto.

Para una determinada pareja de aminoácidos, independientemente de la proximidad del centro de masas, existirán átomos de la cadena lateral del primero en contacto con los de la cadena lateral del otro aminoácido. Por ello, los átomos involucrados se dividen de acuerdo con la siguiente clasificación atómica en 8 clases:

- Hidrofílicos (I). Nitrógenos y oxígenos que pueden donar y aceptar puentes de hidrógeno (ej. Oxígeno de grupos hidroxilo de residuos Ser o Thr)
- Aceptores (II). Nitrógenos u oxígenos que solo pueden aceptar un puente de hidrógeno
- Donador (III). Nitrógenos que solo pueden donar un puente de hidrógeno
- Hidrofóbicos (IV). Todos los átomos de carbono que no están presentes en anillos aromáticos ni están unidos covalentemente a nitrógeno u oxígeno y los halógenos cloro, bromo y yodo
- Aromáticos (V). Átomos de carbono en anillos aromáticos, sin importar cualquier otro tipo de enlace del átomo
- Neutros (VI). Átomos de carbono con enlace covalente a, al menos, un átomo hidrofílico o dos o más átomos aceptores; átomos metálicos; azufre, flúor y fósforo
- Donadores neutros (VII). Átomos de carbono que poseen un enlace covalente con un solo átomo donador

Tipo de contacto	I	II	III	IV	V	VI	VII	VIII
I (Hidrofílico)	+	+	+	-	+	+	+	+
II (Aceptor)	+	-	+	-	+	+	+	-
III (Donador)	+	+	-	-	+	+	-	+
IV (Hidrofóbico)	-	-	-	+	+	+	+	+
V (Aromático)	+	+	+	+	+	+	+	+
VI (Neutral)	+	+	+	+	+	+	+	+
VII (Neutral-donador)	+	+	-	+	+	+	-	+
VIII (Neutral-aceptor)	+	-	+	+	+	+	+	-

Tabla V.16: Tipos de contacto ‘legítimos’ (+) e ‘ilegítimos’ (-), según la metodología CSU [203]

- Aceptores neutros (VIII). Átomos de carbono que poseen un enlace covalente con un solo átomo aceptor

El programa evalúa la naturaleza de los contactos de los diversos átomos de las cadenas laterales que intervienen. Para ello se define el concepto de ‘legitimidad’ de contacto, discriminando aquellos contactos según la clase a la que pertenecen la pareja de átomos involucrada, que mostramos en la Tabla V.16, en contactos ‘legítimos’ e ‘ilegítimos’, según sus propiedades físico-químicas. Al final del proceso, el programa evalúa, con todos los datos disponibles, si el contacto es una verdadera interacción (contacto *nativo*), o simplemente se encuentran próximos en la estructura y no media ninguna interacción entre ellos (contacto *no nativo*). Los contactos considerados *nativos* engloban todos aquellos tipos de interacción vistos en proteínas: contactos hidrofóbicos, puentes de hidrógeno, contactos aromático-aromático, contactos hidrofílico-hidrofóbico, contactos aceptor-aceptor, contactos donador-donador, etc.

Modelo tipo Gō

En nuestro modelo de simulación, los contactos necesarios entre pares de aminoácidos, los mínimos e indispensables para mantener la estructura plegada, serán los contactos *nativos*, y mediará entre ellos un potencial atractivo, mientras que para el resto de posibles contactos, los *no-nativos*, emplearemos un potencial repulsivo simple para evitar el solapamiento entre esferas. Además de este tipo de interacciones, también incluimos interacciones de enlace, angular y de torsión, cuyos valores de equilibrio corresponden a los de la estructura estática. A continuación se detallan las expresiones de estos potenciales.

V.4. SIMULACIÓN DE PROTEÍNAS PARCIALMENTE DESORDENADAS

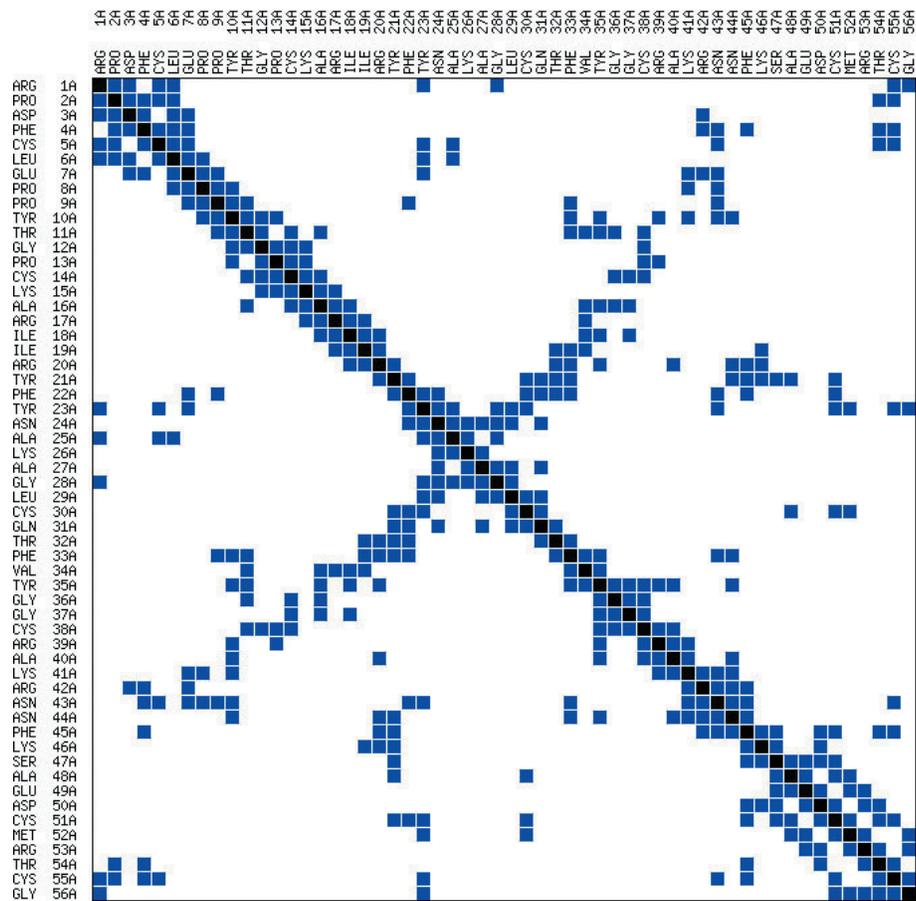


Figura V.11: Mapa de contactos, obtenido mediante el software CSU, de la proteína BPTI (código PDB: 1BPI), producido por la aplicación web alojada en <http://bip.weizmann.ac.il/oca-bin/lpccsu/>.

- *Potencial de enlace.* De nuevo, la distancia entre $C\alpha$ debe mantenerse fija a una distancia de 3.8 Å. Emplearemos un potencial armónico simple (ec. V.5) con una constante de energía relativamente elevada $K_i = 100.0$ Kcal/mol/Å², mientras que la distancia de equilibrio será la que presente en la estructura, muy cercana a los 3.8 Å.
- *Potencial de ángulo de enlace.* Emplearemos un potencial angular armónico simple (ec. V.6) con una constante de energía $Q_i = 20.0$ Kcal/mol/rad². El ángulo de equilibrio interno será el que presente en la estructura.
- *Potencial de ángulo diedro.* Se utiliza un potencial armónico de series de coseno (ec. V.9) con una constante de energía $K_\phi^{(1)} = 1.0$ Kcal/mol y $K_\phi^{(3)} = 0.5$ Kcal/mol. El ángulo de equilibrio será el que presente en la estructura.

$$V_\phi(\phi_i) = K_\phi^{(1)}[1 - \cos(\phi_i - \phi_{0i}^{(1)})] + K_\phi^{(3)}[1 - \cos 3(\phi_i - \phi_{0i}^{(3)})] \quad i = 1, \dots, N-3 \quad (\text{V.9})$$

- *Potencial de elementos no-enlazados.* Como se comenta anteriormente, en el modelo tipo Gō existen dos tipos de interacciones entre elementos no-enlazados: “atractivo-repulsivo” y “puramente repulsivo”, dependiendo de que se traten de contactos nativos o no-nativos, respectivamente. El potencial atractivo-repulsivo es un potencial Lennard-Jones 12-10 modificado (ec. V.10, término V_{ne}^{nativo}), con una constante de energía $\epsilon_1 = 4.0$ Kcal/mol, un σ_{ij} igual a la distancia entre ambos $C\alpha$ en la estructura, y un corte de potencial (*cutoff*) de $3 \times \sigma_{ij}$. Por otro lado, el potencial repulsivo procede de un truncamiento del potencial anterior, conservándose únicamente el término elevado a la decimosegunda potencia (ec. V.11), con una constante de energía $\epsilon_2 = 2.0$ Kcal/mol, un $\sigma_{ij} = 4$ Å, y una distancia de corte del potencial (*cutoff*) de 10 Å. Ambos potenciales serán empleados en dinámica browniana.

$$V_{ne}^{nativo}(i, j) = \epsilon_1 \left[5 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{10} \right], \quad \text{si } r_{ij} \text{ es un contacto nativo} \quad (\text{V.10})$$

$$V_{ne}^{no-nativo}(i, j) = \epsilon_2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12}, \quad \text{si } r_{ij} \text{ es un contacto no-nativo} \quad (\text{V.11})$$

Sin embargo, en el caso de Monte Carlo y MC-Metrópolis, este último potencial repulsivo (ec. V.11) puede ser reemplazado por un potencial de esfera dura discontinuo (ec. V.7), mucho más eficiente computacionalmente, con una distancia límite de 4.5 Å. En la práctica, ambos potenciales repulsivos son equivalentes, como se comentó en la sección V.4.1.

Los valores numéricos que usamos para los parámetros del modelo tipo Gō son los habituales de la bibliografía; ver por ejemplo [65].

Metodologías específicas

El proceso de construcción del modelo se basa en coordenadas atómicas que en su mayor parte pertenecen a estructuras tridimensionales (obtenidas por cristalografía de rayos X o NMR) publicadas en *www.pdb.org*. Sin embargo, en la mayoría de los casos no se dispone de la proteína completa, debido a que generalmente las regiones flexibles no se resuelven, o bien solo se encuentran disponibles los dominios globulares estructurados. Las regiones flexibles no resueltas, como bucles (*loops*), tallos terminales (*tails*) o los segmentos flexibles de unión entre dominios estructurados (*linkers*) son añadidos a la estructura mediante el programa MODELLER[205], muy utilizado en tareas de modelado molecular.

Partiendo de las coordenadas atómicas de la proteína completa, se construyen los archivos donde se indican el número y el tipo de interacciones entre los elementos del modelo, además de la conformación inicial, que servirán de entrada al programa de simulación de dinámica browniana BROWFLEX. Para ello, ha sido necesario construir un programa para tal efecto, que hemos denominado SimuGoLike, que trata de manera diferente las zonas estructuradas y las flexibles:

- **Zonas estructuradas:** En las zonas estructuradas se calculan todas las distancias, ángulos y diedros entre los C α consecutivos que componen la estructura PDB, obviando el resto de átomos. Estas distancias, ángulos y diedros son los parámetros de equilibrio r_0 , α_{i0} y ϕ_{0i} en los potenciales V.5, V.6 y V.9, respectivamente. A continuación, y mediante una subrutina propia con base al software CSU, se diferencian los posibles contactos nativos que se dan entre los pares de elementos, concepto que ya hemos detallado anteriormente. Si se considera un par como contacto nativo (y no media entre ellos ninguna interacción de enlace, ángulo o diedro), se calcula la

distancia entre los $C\alpha$ de ambos aminoácidos, y es tomada como distancia de equilibrio (σ_{ij}) en el potencial definido en la ecuación V.10. Para el resto de posibles pares (de nuevo, no unidos por interacciones de enlace, ángulo o diedro) se emplea un potencial de volumen excluido definido en la ecuación V.11.

- **Zonas flexibles:** En este caso, se asignan los potenciales V.5 y V.6, que mantienen una distancia entre pares de elementos consecutivos de 3.8 Å, y la distribución angular que se halló en la sección V.3.1. Para las interacciones entre pares de elementos no enlazados, se asigna el potencial repulsivo V.8.

Como hemos indicado, la simulación de dinámica browniana se realiza mediante el programa **BROWFLEX**, incluido en el paquete de simulación **SIMUFLEX** [35]. Como se indica en el capítulo II, **BROWFLEX** trabaja sobre un modelo mecánico de esferas generalizado, el cual genera la trayectoria a partir del campo de fuerzas que recopila todas las interacciones. Los tipos y parámetros de las interacciones se indican en los archivos de entrada. **BROWFLEX** es una herramienta muy útil para la simulación de este tipo de proteínas, ya que nos proporciona un conjunto de conformaciones sobre las cuales podemos obtener las propiedades en disolución mediante la metodología del cuerpo rígido. Además, al tratarse de una dinámica browniana con la posibilidad de incluir la interacción hidrodinámica (DB-IH), podemos extraer información de la dinámica traslacional y rotacional, tanto global como interna, de la molécula. Para la generación de trayectorias brownianas del modelo mecánico en disolvente viscoso, incluyendo los efectos de la interacción hidrodinámica, empleamos el algoritmo de Ermak-McCammon [30], con tiempo de paso (*timestep*) de 50 femtosegundos, compatible con el empleado por otros autores [206, 207] para simulaciones de proteínas con el modelo tipo $G\bar{o}$ o similares, recalculando el tensor de difusión cada 10 pasos. Las simulaciones comienzan a partir de periodos de “calentamiento” (que no se registran) equivalentes al 10% de la longitud final de la trayectoria.

Al finalizar la simulación, se procede al cálculo de sus propiedades. Las propiedades en disolución clásicas se calculan a partir de las coordenadas $C\alpha$ mediante el procedimiento de cuerpo rígido (RB) $-D_t(\text{RB})$, $[\eta]$, s , empleando la parametrización hallada en la sección V.2. Se obtienen como el promedio conformacional sobre los valores calculados para las conformaciones individuales, extraídas de la trayectoria. Para obtener una estimación del error, la trayectoria se divide en 4 ó 5 sub-trayectorias, se calculan los valores promedio para cada una de ellas y se halla el error a partir de la desviación estándar de dichos valores. El cálculo de $D_t(CM)$, se realiza a partir de la función de correlación del

desplazamiento cuadrático medio del centro de masas de la molécula (ecuaciones II.11 y II.15), mientras que los coeficientes rotacionales se calculan a partir de su dinámica reorientacional, generalmente a partir de vectores de enlace $C\alpha-C\alpha$ o vectores a lo largo de la molécula o fragmento. Si no se indica lo contrario, los parámetros físico-químicos del disolvente empleados, tanto en la simulación como en el cálculo hidrodinámico, son: temperatura de 20° C, densidad $\rho = 1.0 \text{ g/cm}^3$ y viscosidad $\eta_0 = 0.01$ poises.

Los parámetros de relajación de $^{15}\text{N-NMR}$ T_1, T_2, NOE se calculan a partir de la dinámica reorientacional de los vectores de enlace $C\alpha-C\alpha$, asumiendo que la mayor contribución a la dinámica del enlace N-H es dada por la dinámica del esqueleto peptídico, representada en nuestro modelo por el enlace virtual que conecta dos carbonos α , tal y como proponen Caballero-Manrique *et al* [208]. El procedimiento consiste en hallar los tiempos de correlación τ_i y amplitudes a_i del residuo j a partir de la dinámica del enlace $C\alpha_{j-1}-C\alpha_j$ mediante un ajuste multiexponencial de la función $\langle P_2 \rangle$ (ecuaciones II.34 y II.36). A partir de τ_i y a_i se calcula la función de densidad espectral $J(\omega)$ (ec. II.40), y por último, se obtienen las relaciones T_1/T_2 mediante la ecuación II.50 (más detalles en la sección II.1.2.c). En la relación T_1/T_2 , el efecto del movimiento interno es eliminado, y este depende exclusivamente de los tiempos de relajación anisotrópicos de la difusión rotacional global de la molécula. Sin embargo, se trata de un método aproximado, que no tiene en cuenta las diferentes posiciones y orientaciones de los enlaces N-H en partículas anisotrópicas, al no estar representados en nuestro modelo. Sin embargo, se obtienen resultados aproximados de los ratios T_1/T_2 experimentales a lo largo de la secuencia.

Para el cálculo de curvas de dispersión de rayos X (SAXS) se utiliza el programa CRY SOL [209], a partir de un modelo atómico reconstruido a partir del modelo $C\alpha$, de nuevo como el promedio conformacional sobre las curvas calculadas de las conformaciones individuales, extraídas de la trayectoria. Al final del proceso se obtiene una curva única, de la cual se calcula la distribución de distancias $P(r)$ con el programa GNOM [209]. El proceso de reconstrucción del modelo atómico permite estimar las posiciones del resto de átomos que forman el esqueleto peptídico (grupos NH, CO) y las cadenas laterales, teniendo en cuenta los impedimentos estéricos y las particularidades estructurales de los aminoácidos. Este procedimiento se realiza con el fin de obtener datos que se puedan comparar directamente con las curvas de SAXS experimentales, que al contrario que las propiedades en disolución clásicas, son muy sensibles al detalle estructural “fino”. La reconstrucción se realiza mediante el programa PULCHRA [210].

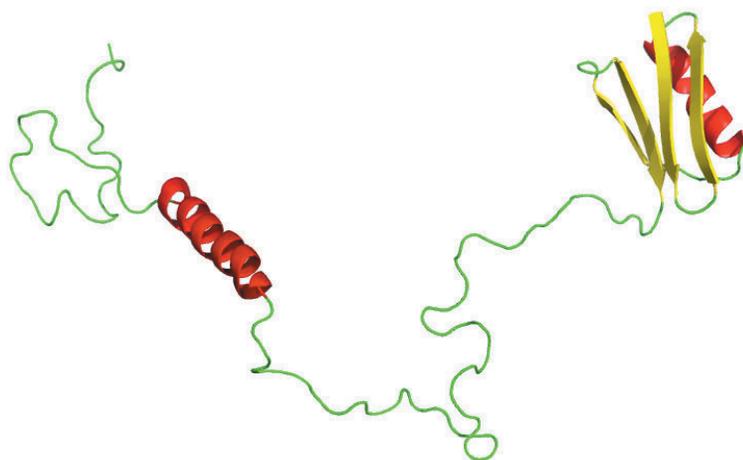


Figura V.12: Representación de la proteína hipotética, quimera 1PGA

V.4.3. Aplicaciones

A continuación, se representan diversos ejemplos de proteínas parcialmente desordenadas que hemos usado para comprobar la validez de nuestro método. En primer lugar, una proteína ficticia, y después pasaremos a aplicar nuestra metodología a casos reales, con el objeto de someterla a comparación con la evidencia experimental.

V.4.3.a. Proteína quimérica 1PGA

A continuación, efectuamos como “prueba de concepto” de nuestra metodología, cálculos para un modelo de proteína hipotética ficticia, de 156 aminoácidos, compuesta por un dominio globular N-terminal, unido a una hélice por un segmento flexible, y por último, un tallo C-terminal flexible, a continuación de la hélice, tal y como podemos observar en la Figura V.12. Los detalles del modelo son los siguientes:

- **Dominio globular N-Terminal:** Dominio de unión de inmunoglobulina de la proteína G de *Streptococcus*. Consta de 56 aminoácidos, y la estructura se caracteriza por presentar una hélice- α rodeada de dos hojas- β antiparalelas. Se encuentra depositado en el *Protein Data Bank* bajo el código 1PGA.
- **Segmento flexible:** Zona flexible de 44 aminoácidos que une el dominio globular C-terminal con la hélice.

- **Hélice:** Zona plegada en forma de hélice- α de 25 aminoácidos. La estructura pertenece a un péptido antimicrobiano, que se encuentra depositado en el *Protein Data Bank* bajo el código 1Z64.
- **Tallo C-terminal,** de 31 aminoácidos, flexible.

Dinámica browniana

Se realizaron cuatro simulaciones simultáneas, de dinámica browniana con interacción hidrodinámica, de un tiempo de 4 μ s, a partir de cuatro conformaciones iniciales diferentes e independientes. El tiempo de cálculo no superó las 24 horas en un equipo Intel Xeon Quad 2.80 GHz, con 8 GB de RAM, corriendo las 4 simulaciones a la vez.

Resultados

Para obtener los resultados se analizaron las 4 trayectorias simultáneamente, para obtener una estimación del error basada en la desviación estándar de cuatro valores. Para el cálculo de propiedades en disolución en equilibrio (R_g , D_t , $[\eta]$) se utilizó el tratamiento RB, cuyos resultados podemos observar en la Figura V.13, junto al coeficiente de difusión traslacional calculado a partir de la dependencia temporal del desplazamiento cuadrático medio del centro de masas de la molécula. Como se menciona anteriormente, este es un procedimiento riguroso para hallar D_t a partir del movimiento browniano, mientras que el tratamiento RB involucra ciertas aproximaciones. No obstante, el coeficiente de difusión obtenido por el tratamiento RB no difiere excesivamente (aprox. un 3%) del calculado mediante el desplazamiento cuadrático medio (7.07×10^{-7} frente a 7.29×10^{-7} cm²/s, respectivamente).

Para estudiar la dinámica rotacional o interna, mucho más informativa que la traslacional, se calculó la función de correlación reorientacional $\langle P_2 \rangle$ (segundo polinomio de Legendre) para diversos vectores de la molécula, tales como el vector extremo extremo, y vectores de enlace C α -C α situados en el glóbulo N-terminal, en la hélice y en el extremo C-terminal (ver Figura V.14).

Por último, en un estudio más detallado de la conformación, obtenemos la intensidad de dispersión de radiación electromagnética $I(h)$ y la distribución de distancias $P(r)$ del modelo atómico reconstruido, que podemos observar en la Figura V.15.

Prop.	Promedio	Error
R_g (Å)	34.5	\pm 0.03
D_t ($\times 10^7$ cm ² /s)	7.07	\pm 0.01
s (S)	2.87	\pm 0.04
$[\eta]$ (cm ³ /g)	16.1	\pm 0.1

Tabla V.17: Propiedades en disolución de quimera 1PGA obtenidas de la simulación

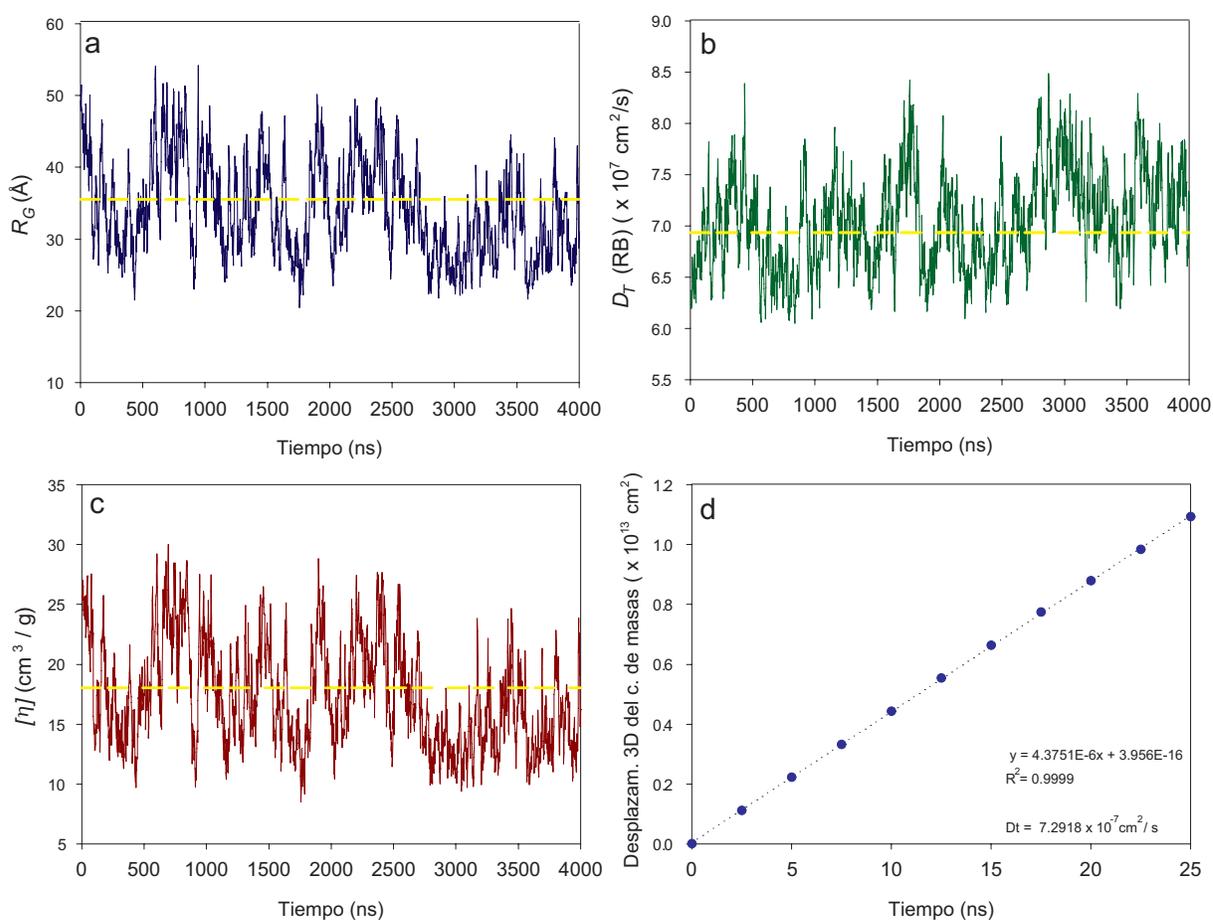


Figura V.13: Resultados obtenidos de la simulación de la proteína quimera 1PGA. Evolución del R_g (a), D_t (b) y $[\eta]$ (c) frente al tiempo, de una de las 4 trayectorias independientes de la simulación. La línea amarilla discontinua es el promedio obtenido de la trayectoria individual, empleando la metodología del cuerpo rígido (RB). (d) Cálculo de D_t mediante el desplazamiento cuadrático medio.

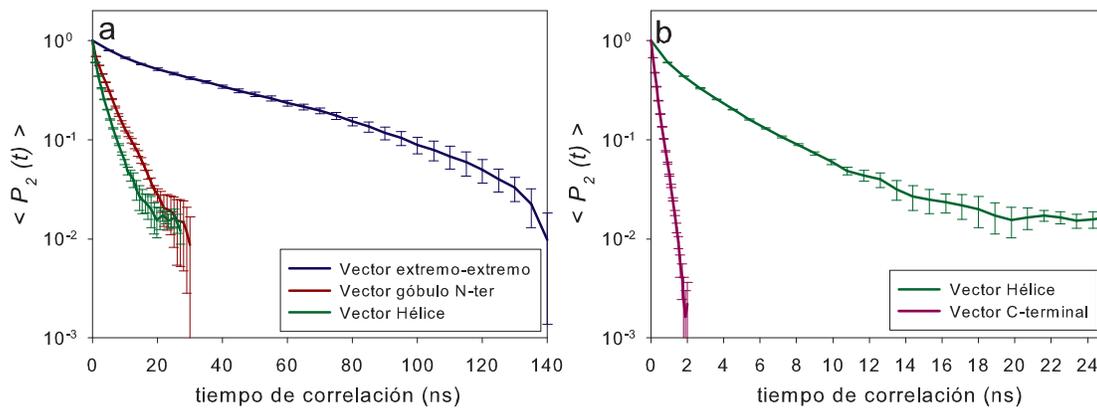


Figura V.14: Funciones de correlación $\langle P_2(t) \rangle$ en escala semilogarítmica de diversos vectores de la molécula. **(a)** Vector extremo-extremo (azul), vector enlace dentro del glóbulo N-terminal (residuos 20-21, línea roja), vector enlace dentro de la hélice (residuos 112-113, línea verde). **(b)** Vector enlace C-terminal (violeta), junto al vector enlace dentro de la hélice (residuos 112-113, línea verde).

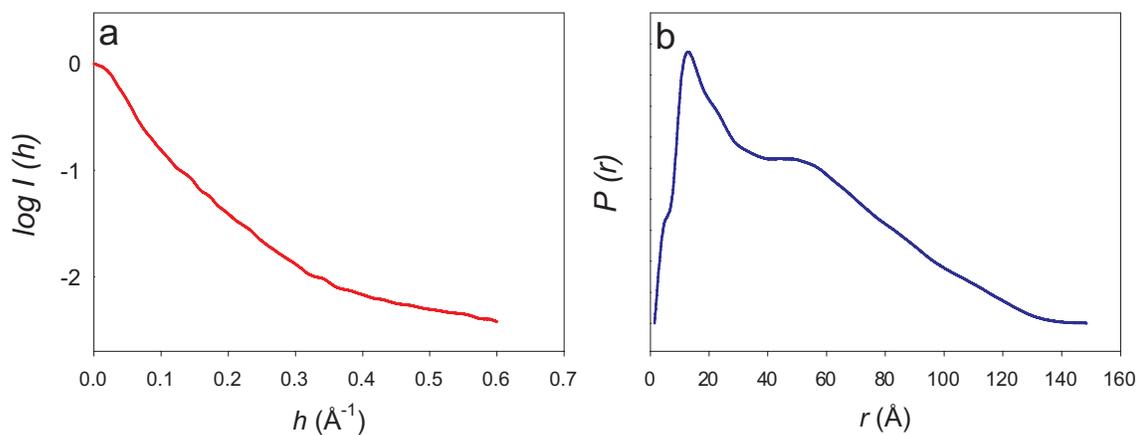


Figura V.15: Funciones de intensidad de dispersión **(a)** y distribución de distancias **(b)**, resultado de promediar las 4 trayectorias de la simulación.

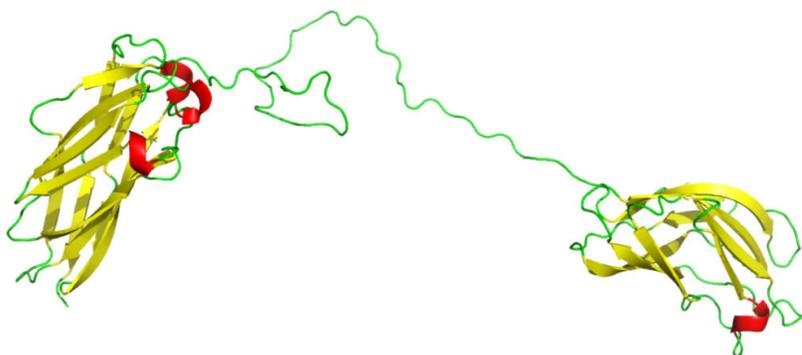


Figura V.16: Representación de la proteína S4

V.4.3.b. Proteína S4

El celulosoma es un gran complejo macromolecular secretado por ciertos microorganismos anaeróbicos para la degradación de la celulosa de origen vegetal. Esta maquinaria enzimática, con un peso molecular del orden de megadaltons, se organiza como un ensamblaje de varias proteínas. La S4 es una construcción quimérica, diseñada para unir dominios catalíticos de diferentes especies, con el fin de obtener una actividad aumentada respecto a la forma silvestre. Esta proteína, de 336 aminoácidos, consiste en dos dominios globulares unidos por un largo segmento desestructurado de 50 aminoácidos [211]. Ambos dominios son muy similares en tamaño, y no se ha descrito ninguna interacción entre ellos.

Modelado y simulación

Siguiendo el mismo procedimiento que en el caso anterior, construimos en primer lugar la proteína bi-globular a partir de las coordenadas atómicas disponibles en el *Protein Data Bank*. Debido a que es una construcción sintética y ambos dominios pertenecen a especies distintas, estos aparecen publicados en entradas diferentes, concretamente, el dominio cohesina de *Clostridium cellulolyticum*, con código 1G1K, y el dominio cohesina de *C. thermocellum*, con código 1OHZ. Posteriormente se añadió el linker flexible de 50 aminoácidos empleando el programa MODELLER.

A continuación, se realizó una única simulación de dinámica browniana con interacción hidrodinámica con un tiempo total de 20 μ s, dividida en 5 sub-muestras, guardando una conformación cada nanosegundo. Se calcularon las curvas de SAXS y la distribución de distancias del modelo atómico reconstruido, que se pueden observar, junto a la expe-

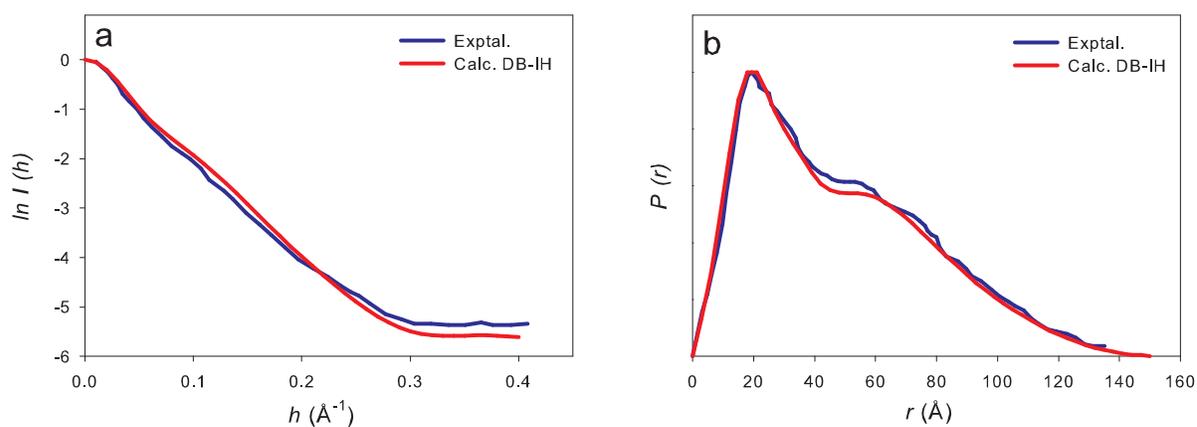
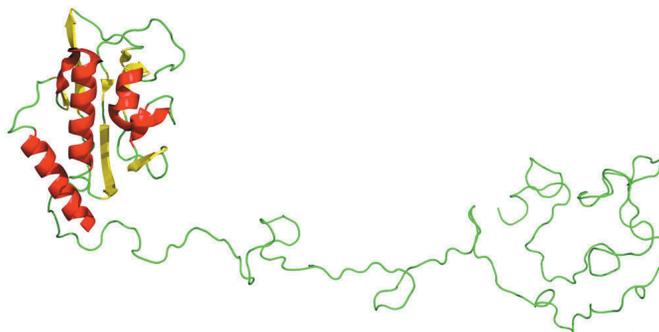


Figura V.17: Curvas de intensidad de dispersión de rayos X (a) y distribución de distancias (b) experimental (línea azul) frente a las calculadas con el modelo atómico (línea roja).

rimental en la Figura V.17. Disponemos de la curva de dispersión de SAXS obtenida de Hammel *et al* [211], que, junto al R_g hallado también por SAXS en el mismo trabajo, son los únicos datos experimentales de que disponemos a día de hoy. Por lo tanto, el análisis de la simulación será exclusivamente de tipo conformacional.

Resultados. Análisis conformacional

El radio de giro calculado de las coordenadas de la trayectoria ($38.3 \pm 2 \text{\AA}$) y el calculado a partir de la curva de SAXS que se obtiene de la simulación (38.9\AA) están muy próximos al experimental ($39.8 \pm 0.2 \text{\AA}$). Además, se observa una buena superposición en las curvas de intensidad de dispersión y en la distribución de distancias, mostradas en la Figura V.17.

Figura V.18: Representación de la proteína *s1ZipA*

V.4.3.c. Proteína ZipA

Estructura de la proteína ZipA

La proteína ZipA es una proteína anclada a la membrana mediante una hélice N-terminal, seguida de un segmento desestructurado y con un dominio globular plegado en el extremo C-terminal. Respecto al segmento flexible, este posee a su vez una zona rica en residuos cargados (zona +/-) y otra zona posterior rica en prolina y glutamina, denominada zona P/Q. Inmediatamente después, comienza el dominio globular plegado, de 140 aminoácidos, del cual se conoce la estructura y se encuentra depositada en el *Protein Data Bank* bajo el código 1F46. Para poder estudiar las propiedades en disolución, el laboratorio del Prof. G. Rivas expresó una variante sin la hélice N-terminal, con el fin de mejorar su solubilidad en medio acuoso. Esta variante, se denomina *mutante s1 soluble* (s1ZipA). Incluyendo la cola N-terminal de histidinas, el segmento desestructurado tiene ahora 183 aminoácidos, y junto al dominio globular C-terminal, presenta un total de 323 residuos. En la Figura V.19 podemos observar una vista esquemática de s1ZipA. Las propiedades en disolución experimentales de que disponemos para s1ZipA son el coeficiente de sedimentación, con un valor de $2.0 S$ ($a_T = 43 \text{ \AA}$) y el radio hidrodinámico a_T , de 35 \AA , obtenido de medidas de difusión, a partir del valor medido en condiciones estándar de $D_t = 6.14 \times 10^{-7} \text{ cm}^2/\text{s}$.

Simulación mediante dinámica browniana

Se lanzaron un total de 64 simulaciones simultáneas de nuestro modelo de ZipA, variando la conformación inicial, que fueron obtenidas mediante el método Monte Carlo-CARP (ver a continuación). En las simulaciones, se incluyó el efecto de la interacción hidro-

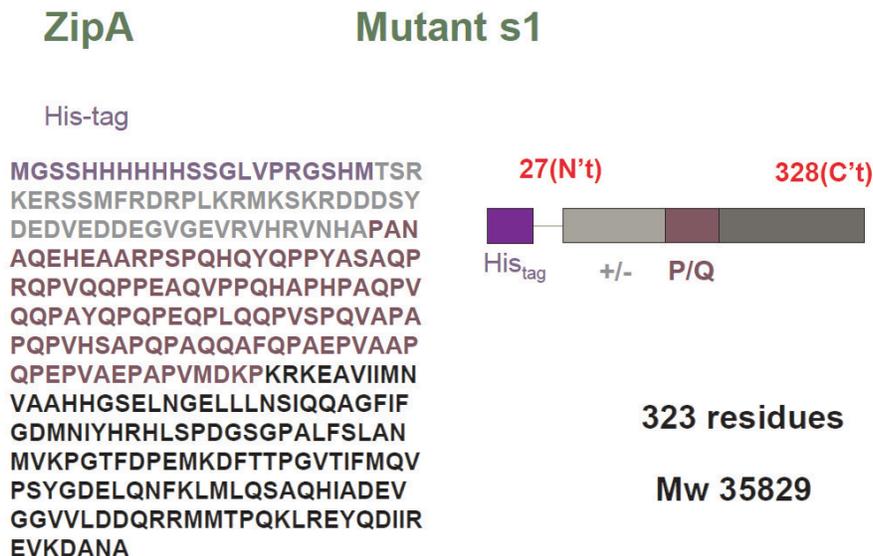


Figura V.19: Secuencia y vista esquemática de la proteína ZipA.

dinámica (IH), y tuvieron una longitud total de 15 μ s cada una. Se obtuvieron un total de 30000 conformaciones en cada simulación y se calcularon las propiedades de interés cada 100 conformaciones, empleando un radio hidrodinámico de los aminoácidos de $r_{HYD} = 6.11$ Å, tal y como se había parametrizado anteriormente (Ver sección V.2). Los resultados se muestran en la Tabla V.18.

Propiedad	Valor	a_X (Å)
R_g (Å)	50.2 ± 6.0	64.85
D_t ($\times 10^7$ cm ² /s)	5.10 ± 0.02	42.07
s (S)	2.04 ± 0.01	42.07
$[\eta]$ (cm ³ /g)	17.93 ± 0.16	46.36

Tabla V.18: Resultados de la simulación Browniana IH.

Simulación mediante MonteCarlo-CARP

Para comprobar la equivalencia de los métodos Monte Carlo y la dinámica browniana, se realizó en este caso además un cálculo Monte Carlo-CARP, cuyo procedimiento detallamos a continuación. Partimos de las coordenadas de los C α del dominio globular plegado (140 elementos), a continuación, y a una distancia exacta de 3.8 Å del primer elemento del dominio estructurado, se coloca al azar el primer elemento de la cadena flexible, y a

partir del mismo, se construye la cadena flexible (183 elementos) hasta alcanzar el total de 323 elementos. Además, se añadieron restricciones de volumen excluido, tanto para la cadena flexible consigo misma, como para la cadena frente al dominio globular estático. Se obtuvieron las propiedades a partir de 2000 conformaciones, asignando un radio de volumen excluido $r_{HS} = 2.25 \text{ \AA}$. Una vez obtenidas las conformaciones, se calcularon las propiedades hidrodinámicas asignando distintos radios hidrodinámicos r_{HYD} , desde 4.5 hasta 6.5 \AA , en intervalos de 0.5 \AA . Los resultados se muestran en la tabla V.19.

r_{HYD} (\AA)	Prop.	Value	a_X (\AA)
4.5	R_g (\AA)	51.1 ± 0.2	66.0
	D_t ($\times 10^7 \text{ cm}^2/\text{s}$)	5.72 ± 0.01	37.6
	s (S)	2.19 ± 0.003	37.6
	$[\eta]$ (cm^3/g)	12.12 ± 0.09	41.0
5.0	R_g (\AA)	51.2 ± 0.2	66.1
	D_t ($\times 10^7 \text{ cm}^2/\text{s}$)	5.12 ± 0.01	42.0
	s (S)	1.96 ± 0.003	42.0
	$[\eta]$ (cm^3/g)	18.73 ± 0.12	47.4
5.5	R_g (\AA)	51.2 ± 0.2	66.1
	D_t ($\times 10^7 \text{ cm}^2/\text{s}$)	5.03 ± 0.01	42.7
	s (S)	1.92 ± 0.003	42.7
	$[\eta]$ (cm^3/g)	20.45 ± 0.13	48.8
6.0	R_g (\AA)	51.3 ± 0.2	66.2
	D_t ($\times 10^7 \text{ cm}^2/\text{s}$)	4.95 ± 0.01	43.4
	s (S)	1.89 ± 0.003	43.4
	$[\eta]$ (cm^3/g)	22.48 ± 0.14	50.4
6.5	R_g (\AA)	51.2 ± 0.2	66.0
	D_t ($\times 10^7 \text{ cm}^2/\text{s}$)	4.88 ± 0.01	44.0
	s (S)	1.87 ± 0.003	44.0
	$[\eta]$ (cm^3/g)	24.63 ± 0.15	51.9

Tabla V.19: Propiedades calculadas mediante la simulación MonteCarlo-CARP con diferentes valores del radio hidrodinámico r_{HYD} .

Conclusiones

En primer lugar, observamos que se obtienen resultados muy similares con un método u otro (ver Tabla V.18 frente a Tabla V.19 para 6.0 \AA), siendo el método Monte Carlo mucho más eficaz computacionalmente. Las propiedades en disolución calculadas por ambos métodos, son compatibles con las halladas experimentalmente.

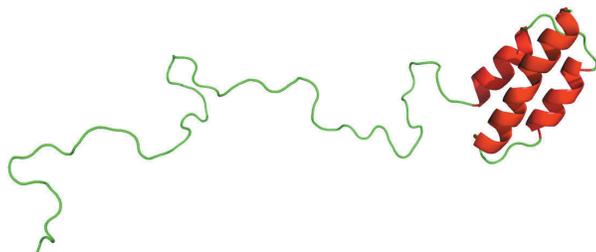


Figura V.20: Representación de la proteína X del virus Sendai.

V.4.3.d. Proteína X del virus Sendai

La RNA-polimerasa (dependiente de RNA) del virus Sendai está compuesta por una proteína con actividad polimerasa denominada L y una fosfoproteína denominada P. Esta última juega un papel importante en la enzima posicionando L dentro del complejo de transcripción N-RNA, formado por el RNA unido a una nucleoproteína denominada N. La fosfoproteína P está compuesta por 568 aminoácidos, y tiene una estructura modular. La parte N-terminal (1-319) actúa como chaperona (asiste al plegamiento) de la nucleoproteína N, además de otras funciones. En la parte C-terminal se han identificado varias regiones discretas que interactúan con otras proteínas en el proceso de la transcripción, estas regiones son: el dominio de unión a L (411-445), que forma parte del dominio de oligomerización (denominado PMD, residuos 320-446), y el dominio de unión a N-RNA (474-568), denominado Proteína X (PX). PX juega un papel crucial en la síntesis de RNA y es responsable directa de la unión al complejo N-RNA. Más detalles acerca de estos aspectos pueden encontrarse en la referencia [212].

Estructura de la Proteína X

La PX posee dos regiones claramente diferenciadas desde el punto de vista estructural. El dominio N-terminal (474-515, 42 residuos) no tiene estructura y es considerado un dominio desordenado. Por otro lado, el dominio C-terminal (516-568, 53 residuos) presenta una estructura secundaria definida, resuelta por NMR y depositada en el *Protein Data Bank* bajo el código 1R4G. Está compuesta por tres hélices- α dispuestas alrededor de un núcleo hidrofóbico.

La PX fue clonada y purificada por Blanchard *et al* [212], incluyéndose una cola de histidinas en el extremo N-terminal de 15 residuos de longitud, por lo que la proteína tiene una longitud total de 109 aminoácidos (56 en el tallo N-terminal y 53 en el glóbulo C-

terminal). Los datos experimentales de que disponemos de esta construcción son: radio de giro, obtenido de experimentos de dispersión de rayos X (SAXS) y dispersión de neutrones (SANS) [212, 213], además de la curva de intensidad de dispersión de rayos X $I(h)$ y el factor de forma $h^2I(h)$ y distribución de distancias $P(r)$. Además de disponer de esta información de tipo conformacional, también disponemos de datos experimentales acerca de la dinámica rotacional interna de la molécula [214], por medio de los tiempos de correlación rotacional τ_c de la región N-terminal por un lado, y la región C-terminal por otro. Estos τ_c , obtenidos de experimentos de NMR, son estimados a partir de los ratios longitudinal y transversal de relajación del espín del ^{15}N (T_1 , T_2), asumiendo que el movimiento interno es mucho más rápido que la difusión rotacional global de la proteína.

Modelado y simulación

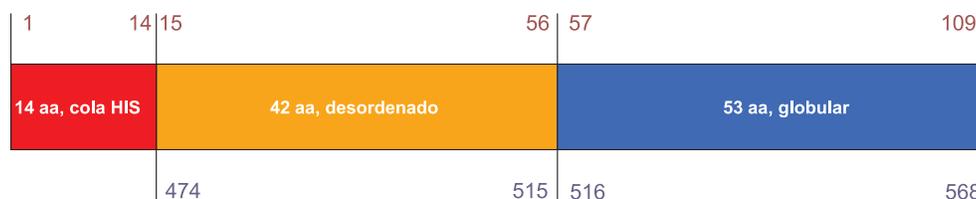


Figura V.21: Esquema de la proteína X del virus Sendai. Los números que aparecen en la parte inferior (azules) corresponden a las posiciones dentro de la secuencia de la proteína P, mientras que los números superiores (rojos) corresponden a la numeración empleada en nuestro modelo

Siguiendo con el procedimiento explicado para los casos anteriores, modelamos la molécula como un modelo reducido que únicamente representa los $C\alpha$ de la proteína, según el esquema que aparece en la Figura V.21. Para el dominio N-terminal (dominio desplegado, aminoácidos 1-56) empleamos el modelo de proteína desnaturalizada desarrollado en esta Tesis. Para el dominio C-terminal (dominio plegado, aminoácidos 57-109), empleamos el modelo tipo $G\bar{o}$, obteniendo la información necesaria de la estructura tridimensional (código PDB 1R4G).

Respecto a los detalles de la simulación, empleando el programa SIMUFLEX se realizó una dinámica browniana con interacción hidrodinámica, dividida en 4 sub-simulaciones de $10 \mu s$ cada una (con un tiempo de calentamiento de 10 ns), partiendo de conformaciones iniciales independientes. Estas conformaciones se construyeron a partir de la estructura tridimensional publicada, añadiéndole el segmento flexible mediante MODELLER. Las cuatro simulaciones fueron lanzadas simultáneamente y tuvieron una duración de 5 días (Equipo

Prop.	Valor
R_g (Å)	26.7 ± 0.1
D_t ($\times 10^7$ cm ² /s)	8.53 ± 0.1
s (S)	1.13 ± 0.01
$[\eta]$ (cm ³ /g)	15.17 ± 0.1

Tabla V.20: Resultados de la simulación Browniana IH para la proteína PX.

linux Xeon Quad 2.6 Ghz). Se almacenaron 20000 conformaciones por sub-simulación (2 conformaciones / ns), que se emplearon posteriormente en su totalidad (80000 conf.) para el cálculo de propiedades, tiempos de correlación y curvas de dispersión de rayos X. Las propiedades y los tiempos de correlación fueron calculados con el programa ANAFLEX, mientras que las curvas de dispersión, factor de forma y distribución de distancias fueron calculadas empleando un programa propio para el modelo reducido de $C\alpha$, que preparamos para tal fin (tiene en cuenta el solapamiento de los elementos del modelo). Se utilizaron los programas CRY SOL/GNOM [209] para el cálculo de estas curvas en el modelo con detalle atómico.

Resultados. Propiedades conformacionales

El radio de giro obtenido de la simulación $R_g = 26.7 \pm 0.1$ Å, calculado a partir de las coordenadas del modelo, está en concordancia con los valores experimentales obtenidos de experimentos de SAXS (29.7 Å) [213] y SANS (25.0 ± 1 Å) [212]. Respecto a las curvas de intensidad de dispersión, factor de forma y distribución de distancias, que podemos ver en la Figura V.22, se observa una buena superposición a la curva experimental, especialmente la correspondiente al modelo atómico. Aunque las curvas calculadas a partir del modelo reducido también se superponen relativamente bien a la experimental, se consiguen mejores resultados si se reconstruye el modelo reducido y se calculan las curvas a partir del modelo atómico.

Resultados. Análisis de la dinámica reorientacional

Mediante funciones de correlación $\langle P_2(t) \rangle$ podemos analizar la dinámica reorientacional global e interna de la molécula. Se analizó la reorientación de varios vectores a lo largo de la molécula, cuyos resultados se muestran en la Figura V.23 y Tabla V.21. Además del vector extremo-extremo, del cual extraemos información de la dinámica global, se analizó el vector extremo-extremo del glóbulo, con el fin de compararlo con el obtenido de

Vector	τ_{arm} (ns)	τ_1 (ns)	τ_2 (ns)
1-109	21.73	-	-
1-56	5.5	3.2 (0.47)	15.5 (0.53)
1-2	0.28	-	-
25-26	0.49	-	-
56-57	0.94	-	-
57-106	7.67	-	-
85-86	0.77	0.59 (0.75)	7.64 (0.25)
58-59	0.97	0.71 (0.71)	8.06 (0.29)
HYDROPRO (PDB)	4.9	-	-

Tabla V.21: Resultados del análisis de la función de correlación $\langle P_2 \rangle$ para la proteína PX, donde aparecen los tiempos de relajación obtenidos por ajustes mono o bi-exponenciales. Entre paréntesis aparece la amplitud normalizada en el caso de ajustes bi-exponenciales.

experimentos de NMR. Además, se analizaron varios vectores de enlace $C\alpha-C\alpha$ a lo largo de la molécula, tanto en el dominio flexible como en el globular.

Analizando la reorientación del vector extremo-extremo del glóbulo, obtenemos un tiempo de correlación $\tau_c = 7.67$ ns, muy próximo al observado experimentalmente (7.3 ns) [214]. Además, mediante el programa HYDROPRO, se hallaron los tiempos de relajación (metodología RB) del glóbulo sin el segmento flexible asociado, obteniendo un $\tau_{arm} = 4.9$ ns. Como era de esperar, la presencia del segmento flexible provoca que la velocidad con la que se reorienta se vea disminuida. Los vectores de enlace $C\alpha-C\alpha$ situados en el glóbulo presentan dos comportamientos diferenciados, requiriendo un ajuste bi-exponencial. Por ejemplo, se obtuvieron para el enlace 85-86 dos tiempos de correlación. El primero de ellos, $\tau_1 = 7.64$ ns ($A_1 = 0.25$) corresponde a la reorientación del glóbulo hallado anteriormente. El otro tiempo obtenido, $\tau_2 = 0.59$ ns ($A_2 = 0.75$) refleja la dinámica local del propio enlace. Se obtienen resultados similares para el otro vector de enlace dentro del glóbulo analizado (58-59).

Respecto al segmento flexible, los vectores individuales de enlace evidentemente deben presentar una dinámica muy rápida debido a su alta flexibilidad. Se analizaron 3 vectores de enlace $C\alpha-C\alpha$, uno situado en el extremo N-terminal (primer enlace 1-2), otro en una región intermedia (25-26) y el último enlace del segmento flexible (56-57), muy próximo al dominio C-terminal globular. Se obtuvieron unos tiempos de correlación de 0.28, 0.49 y 0.94 ns respectivamente. Se observa, como era de esperar, que la dinámica es más lenta a medida que el enlace tiene más limitado su movimiento.

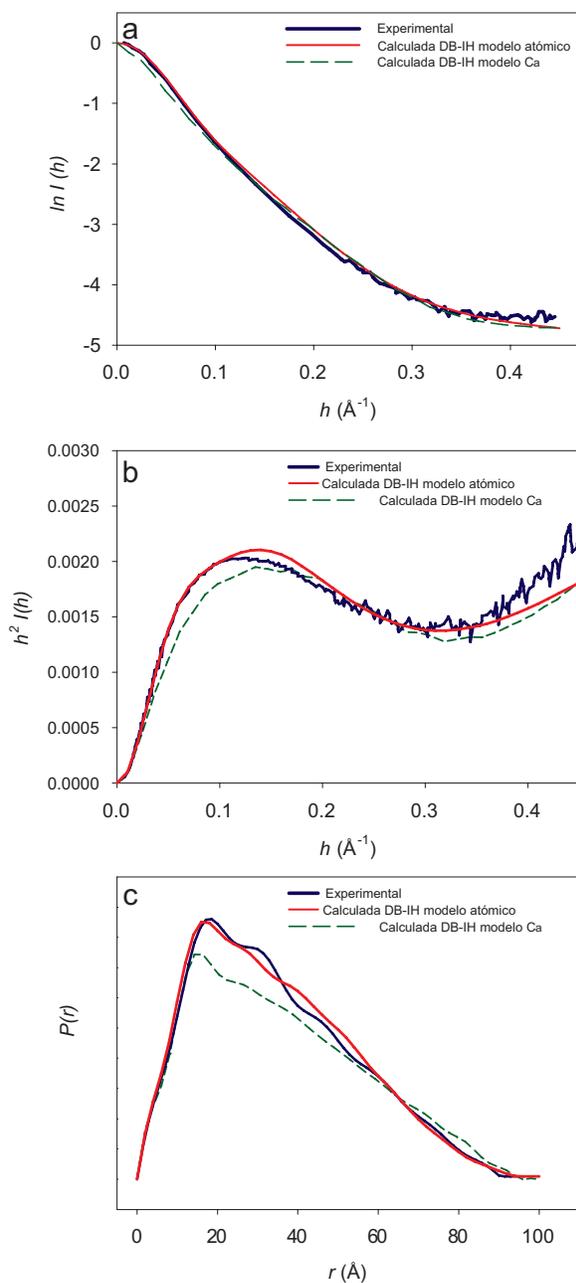


Figura V.22: Curvas de intensidad de dispersión (a), factor de forma (b), y distribución de distancias (c) experimental (línea azul) frente a las calculadas con el modelo $C\alpha$ (línea verde) y con el modelo atómico (línea roja), para la proteína PX.

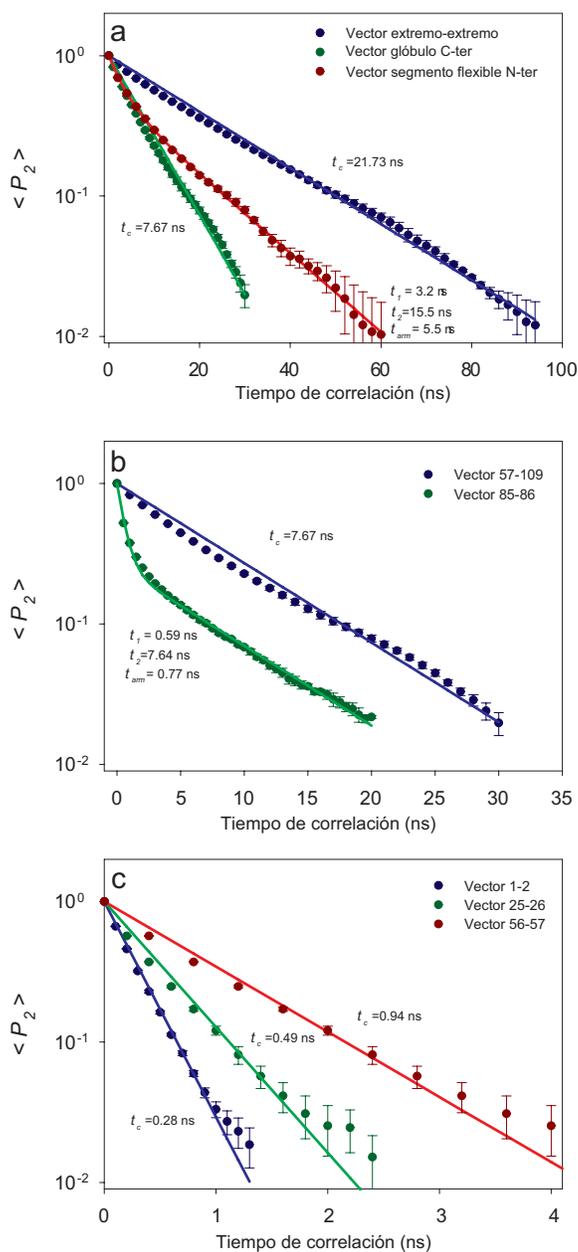


Figura V.23: Funciones $\langle P_2 \rangle$ de correlación para diversos vectores de la proteína PX. **(a)** Representación del vector extremo-extremo (1-109) de la proteína completa, y un vector del glóbulo (57-106) y del segmento flexible (15-45). **(b)** Representación del vector del glóbulo (57-106) junto a un vector $C\alpha-C\alpha$ de enlace situado en su interior (85-86). **(c)** Representación de 3 vectores $C\alpha-C\alpha$ de enlace del segmento flexible.

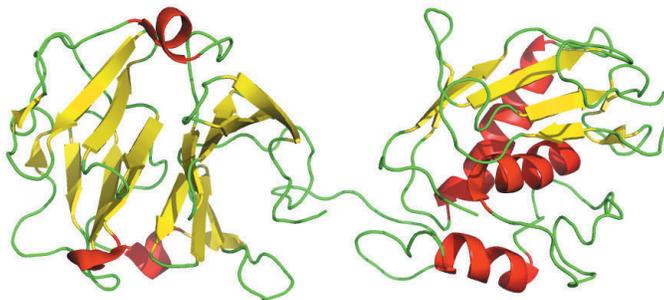


Figura V.24: Representación tridimensional de la proteína MMP-12

V.4.3.e. Proteína MMP-12

Las metaloproteasas de matriz (MMP) son una importante familia de proteínas involucradas en un gran número de procesos extracelulares, incluyendo la degradación de la matriz extracelular, como el colágeno, elastina, proteoglicanos y otros. Casi todas las proteínas de esta familia están compuestas por dos dominios globulares unidos por linkers, siendo uno de ellos (dominio CAT, 157 residuos, en el caso de MMP-12) el que posee la actividad proteasa, mientras se especula con la función del otro (dominio HPX, 184 residuos, en el caso de MMP-12). Algunos autores proponen que puede ayudar a desenrollar la triple hélice de colágeno para facilitar su degradación [215]. La MMP-12 tiene un total de 365 aminoácidos, siendo su linker de pequeña longitud (15 aminoácidos).

Modelado y simulación

En este caso, disponemos de la estructura completa, incluyendo el linker central, en el *Protein Data Bank* bajo el código 3BA0. Debido a la pequeña longitud del linker, suponemos una variabilidad conformacional relativamente pequeña, por lo que una única simulación DB-IH de $10 \mu\text{s}$ es suficiente. En este caso, se guardaron 100000 conformaciones a lo largo de la simulación. Posteriormente, una de cada diez conformaciones guardadas (10000) fueron reconstruidas a nivel atómico y se calcularon las curvas de SAXS.

Resultados. Análisis conformacional

De nuevo, el radio de giro calculado de las coordenadas de la trayectoria ($30.3 \pm 0.7 \text{ \AA}$) y el calculado a partir de la curva de SAXS que se obtiene de la simulación (31.3 \AA)

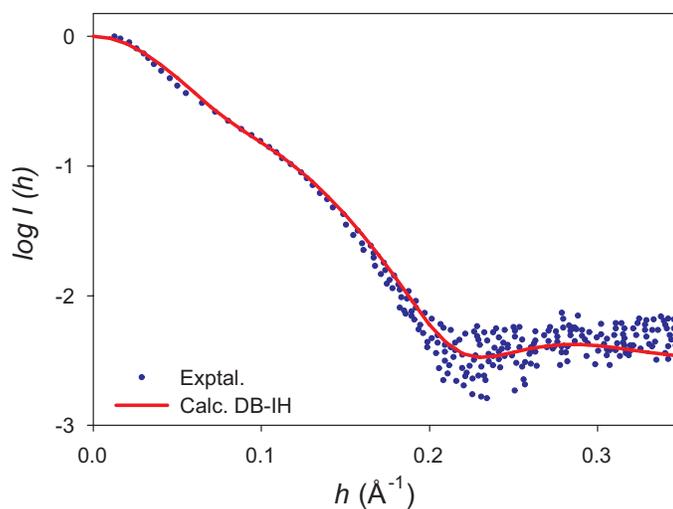


Figura V.25: Curvas de intensidad de dispersión experimental de la proteína MMP-12 (línea azul) frente a la calculada con el modelo atómico (línea roja).

están muy próximos al experimental ($31 \pm 1 \text{ \AA}$) [215]. Además, se observa una buena superposición en la curva de intensidad de dispersión mostrada en la Figura V.25.

Resultados. Análisis relajación NMR

En el cálculo de los parámetros de relajación de NMR, se utilizaron las 100000 conformaciones para hallar los τ y amplitudes. La simulación se hizo a 25°C , y los datos se calcularon para un campo magnético de 700 Mhz (17.6 T) siendo estas las mismas condiciones que en los datos experimentales, obtenidos de Bertini *et al* [215]. En la Figura V.26 se muestran estos parámetros calculados frente a los experimentales.

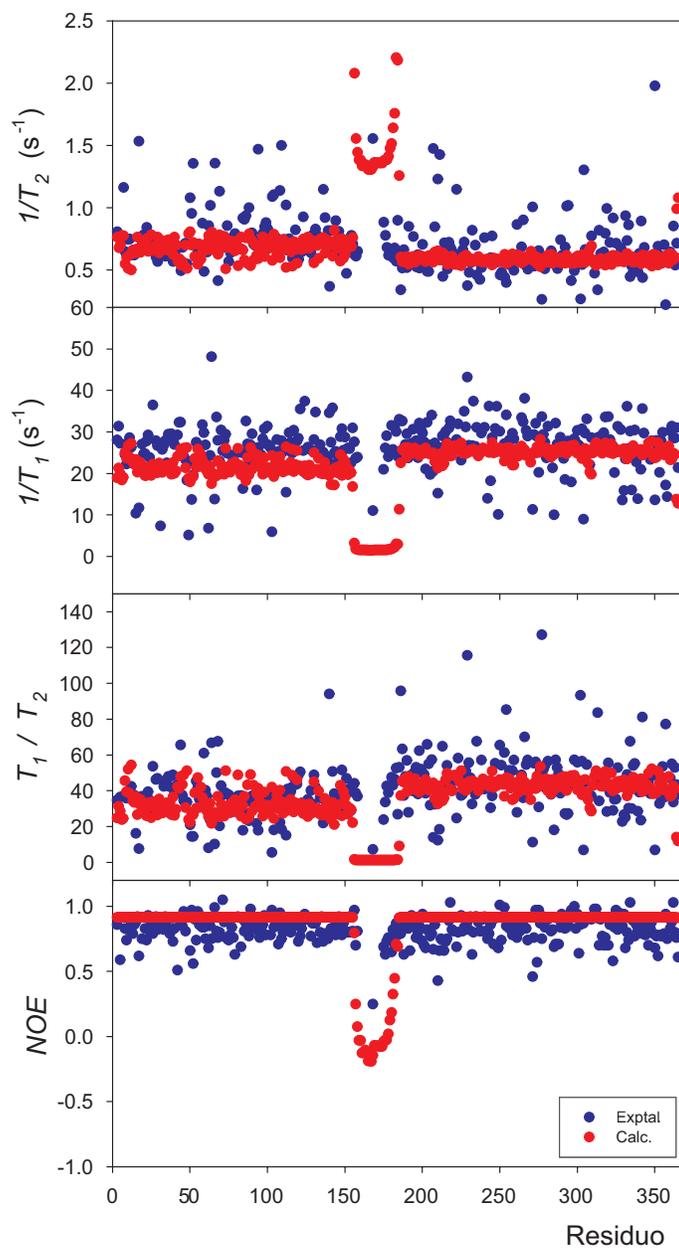


Figura V.26: Análisis de los parámetros de NMR de la proteína MMP-12. Datos experimentales (puntos azules) frente a los calculados (puntos rojos).

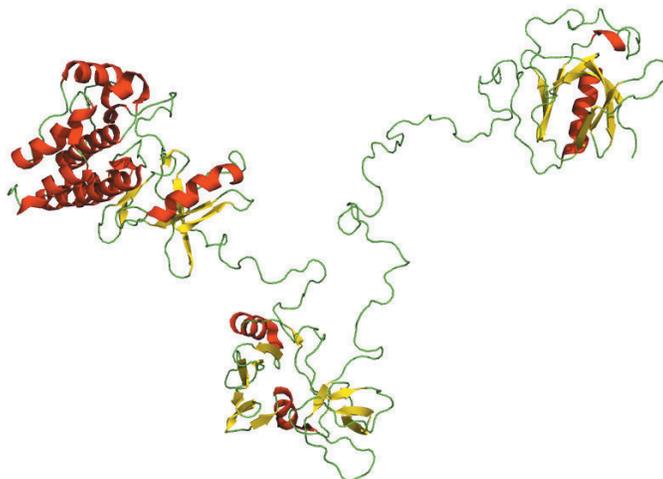


Figura V.27: Organización globular de la proteína BTK

V.4.3.f. Proteína BTK

La Bruton Tirosina Kinasa es una proteína crucial para el desarrollo de los linfocitos B en humano y ratón. Compuesta por 659 aminoácidos, presenta 4 dominios globulares, desde el extremo N-terminal al C-terminal: PH, (169 aa), SH3 (55 aa), SH2 (96 aa), y kinasa (KIN, 258 aa), unidos por 3 linkers de 50, 7 y 21 aminoácidos respectivamente. Los datos experimentales que disponemos son: el R_g experimental, con un valor de 50 ± 0.4 Å [216, 217] y el coeficiente de sedimentación $s_{20,w}$, con un valor de 3.3 S [216].

Modelado y simulación

La estructura completa no se encuentra disponible, por lo que se ha construido a partir de los dominios aislados, para los que sí disponemos de sus coordenadas atómicas en el *Protein Data Bank*. En concreto, se emplearon las estructuras 1BTK, 1AWX, 2GE9 y 1K2P para los dominios PH, SH3, SH2 y KIN respectivamente. Posteriormente se añadieron los linkers empleando el programa MODELLER [205]. Debido al tamaño de esta proteína, y su elevada variabilidad conformacional, se realizaron 6 simulaciones DB sin incluir interacción hidrodinámica, de $10 \mu\text{s}$ cada una, con conformaciones iniciales independientes y guardando una conformación cada nanosegundo. Los datos calculados se obtuvieron del promedio de las 6 simulaciones (60000 conf.).

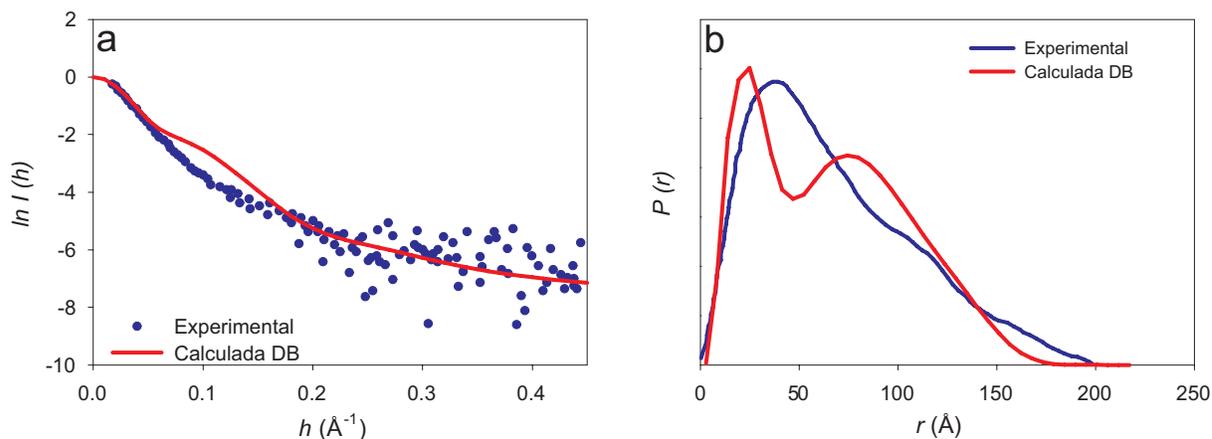


Figura V.28: Curvas de intensidad de dispersión de rayos X (a) y distribución de distancias (b) experimental (puntos/línea azul) frente a las calculadas con el modelo atómico (línea roja).

Resultados. Análisis conformacional e hidrodinámico

Aunque el radio de giro calculado (R_g calc. = $48.2 \pm 1 \text{ \AA}$) coincide con el experimental, observamos importantes diferencias en las curvas de intensidad de dispersión de rayos X y en la distribución de distancias, como podemos observar en la figura V.28. En esta última curva, vemos cómo la distribución calculada presenta dos picos diferenciados frente a un pico único en la experimental. Estas diferencias pueden ser atribuidas a posibles interacciones entre dominios [217], que favorecen una distribución conformacional más compacta, y/o formaciones transitorias de estructuras secundarias en los segmentos flexibles, concretamente en el más largo, donde la alta frecuencia de residuos de prolina pueden dar lugar a formaciones transitorias de hélices de poliprolina [217]. Por otro lado, el coeficiente de sedimentación calculado ($s_{20,w} = 3.92 \pm 0.03 \text{ S}$) se encuentra por encima, aunque muy próximo al experimental 3.3 S [216], debido probablemente a las diferencias conformacionales anteriormente descritas.

V.5. Conclusiones

A través de la metodología desarrollada en este capítulo, somos capaces de reproducir con notable precisión las propiedades de transporte (coeficiente de difusión y sedimen-

tación, viscosidad intrínseca y dinámica rotacional), y propiedades conformacionales de diversos tipos de proteínas.

En primer lugar, en la sección V.2, se han investigado diferentes procedimientos para el cálculo de propiedades para proteínas globulares rígidas de estructura conocida, basados en modelos de esferas, desde resolución atómica (una esfera por átomo) hasta modelos $C\alpha$ de "grano grueso", hallando el radio hidrodinámico óptimo para la esfera que representa el átomo o el aminoácido, dependiendo del caso. Mediante esta parametrización conseguimos reproducir las propiedades experimentales con una precisión muy alta.

A continuación, desarrollamos un modelo $C\alpha$ de "grano grueso" para proteínas flexibles, (sección V.3), basado en estadísticas obtenidas de zonas sin estructura del *Protein Data Bank*, datos experimentales provenientes de proteínas desnaturalizadas y empleando el radio hidrodinámico obtenido de la sección anterior. Mediante este modelo conseguimos una buena reproducción de las propiedades en disolución, dentro de las limitaciones del modelo. Además, observamos que las proteínas completamente desordenadas en condiciones fisiológicas no quedan fuera de la aplicabilidad de este modelo.

Y por último, en la sección V.4, aplicando los avances de las dos secciones anteriores, desarrollamos una metodología para la simulación de proteínas parcialmente flexibles de naturaleza mixta (parte globular rígida, parte flexible) mediante un modelo $C\alpha$. Este modelo, de momento, está limitado a casos concretos, donde las estructuras tridimensionales de las zonas rígidas son conocidas, y además no existen interacciones entre los diferentes dominios de la proteína. En esos casos, conseguimos una buena concordancia entre los datos experimentales y los obtenidos de la simulación. Esta metodología, que destaca por su sencillez, posee numerosas ventajas: Podemos extraer mucha información de una única simulación (propiedades conformacionales, como R_g , curvas de SAXS, e hidrodinámicas, como coeficientes de difusión, sedimentación y propiedades rotacionales, entre otras), que pueden servir para contrastar datos experimentales obtenidos de numerosas técnicas. Además, destaca la rapidez computacional del método. Se pueden simular proteínas relativamente grandes en trayectorias de decenas de microsegundos en apenas unos días de cálculo en equipos de escritorio actuales. Sin embargo, esta metodología todavía presenta aspectos mejorables, como la posibilidad de incluir interacciones entre dominios, estructuras transitorias en los segmentos flexibles, o especificidad de secuencia en estos segmentos (mediante interacciones angulares y de torsión), especialmente en el caso de los aminoácidos glicina y prolina, que daría un mayor realismo a la dinámica de los segmentos flexibles.



Conclusiones

1. Se ha conseguido mejorar y extender la metodología existente (teoría y métodos) para el cálculo o predicción de propiedades en disolución de macromoléculas, y para el análisis de los datos experimentales de estas propiedades en relación con la estructura o conformación que presentan en disolución. Como avances computacionales, hemos mejorado la programación, la gestión de memoria, e implementado la paralelización para máquinas multi-core en el cálculo hidrodinámico y en los programas de simulación molecular, llegando a acelerar los cálculos hasta un factor de veinte respecto a los programas anteriores. Gracias a este avance, hemos sido capaces de estudiar sistemas más grandes y complejos, como grandes (bio)polímeros y proteínas de alto peso molecular, incluyendo el ADN y proteínas multi-globulares.
2. Se ha diseñado un protocolo para el ajuste de valores calculados frente a valores teóricos, simultáneo para varias propiedades, destinado a la búsqueda y refinamiento estructural de (bio) macromoléculas. Como ejemplo de aplicación, se ha determinado la conformación en disolución de moléculas de anticuerpo IgG. Tras construir el modelo, se realizó un barrido conformacional, y se calcularon las propiedades en disolución para cada conformación, para posteriormente hallar la conformación más probable a partir de datos experimentales y un nuevo programa denominado HYDROFIT.

3. Se han logrado avances en el modelo vermiforme para polímeros lineales, de especial relevancia en biopolímeros (ADN, polisacáridos y polímeros sintéticos). Se llevó a cabo un barrido conformacional mediante el método de Monte Carlo, aprovechando los avances computacionales antes detallados, variando únicamente dos parámetros del modelo (longitud de contorno y longitud de persistencia). Tras el cálculo promediado de sus propiedades en disolución (coeficiente de difusión, sedimentación y viscosidad), y conformacionales (radio de giro), se obtuvieron una serie de datos que se emplean posteriormente para estimar parámetros estructurales de especial interés, tales como el diámetro del polímero, la masa por unidad de longitud y la longitud de persistencia. A tal efecto hemos diseñado un protocolo notablemente sistemático y eficaz.
4. Para proteínas globulares rígidas, se han investigado diferentes procedimientos para el cálculo de propiedades de transporte (coeficiente de difusión, sedimentación, viscosidad y rotación) y conformacionales (radio de giro y curvas de dispersión de rayos X), basados en modelos de esferas, desde resolución atómica (una esfera por átomo) hasta modelos de "grano grueso", donde una esfera representa un aminoácido (modelo $C\alpha$). Se han comparado a su vez modelos "macizos", donde se representan todos los átomos o aminoácidos, y modelos de "concha", donde solo se describe la superficie de la macromolécula que está en contacto con el disolvente. Se compararon propiedades calculadas frente a propiedades experimentales, de un amplio conjunto de proteínas, con estructura atómica disponible, con un rango de peso molecular desde aproximadamente 6 kDa hasta 2500 kDa, llegando a la conclusión que los modelos $C\alpha$ predicen las propiedades en disolución con la misma calidad que los modelos de resolución atómica, siendo indiferente emplear modelos "macizos" o "de concha". Estos últimos son los más eficaces computacionalmente. Para ello, los modelos fueron parametrizados variando únicamente el radio hidrodinámico que se asigna a los átomos o aminoácidos, obteniendo radios diferentes en función del modelo empleado. Se ha construido un nuevo programa HYDROPRO para el cálculo de dichas propiedades en disolución, incluyendo los avances metodológicos y computacionales comentados en el punto 1, y la posibilidad de elección del tipo de modelo "macizos" o "de concha", además de una interfaz gráfica.

5. Para proteínas desnaturalizadas, se ha desarrollado un modelo de esferas de “grano grueso”, donde una esfera representa un aminoácido (modelo $C\alpha$), capaz de predecir correctamente propiedades en disolución (radio de giro, coeficiente de difusión, sedimentación y viscosidad) de un amplio conjunto de datos experimentales obtenidos de la literatura, de proteínas en presencia de agentes desnaturalizantes y reductores. Para la parametrización del modelo se empleó el radio hidrodinámico de las esferas hallado en proteínas globulares para modelos $C\alpha$, siendo el único parámetro ajustable el radio de volumen excluido entre las esferas que componen el modelo. Debido a que se trata de un modelo flexible, recurrimos a un barrido conformacional mediante dinámica browniana, Monte Carlo y Metrópolis, obteniendo resultados correctos e idénticos en los tres casos. Además, estudiamos la aplicabilidad del modelo en proteínas intrínsecamente desordenadas, que carecen de estructura en condiciones fisiológicas.

6. Aprovechando la metodología y los avances descritos en los puntos 1, 4 y 5, se ha desarrollado un modelo $C\alpha$ de simulación de proteínas de naturaleza mixta (parte globular rígida, parte flexible), que presentan regiones flexibles (tallos o grandes bucles), y proteínas multi-globulares unidas por segmentos flexibles (*linkers*). Mediante dinámica browniana, reproducimos con notable precisión las propiedades de transporte (coeficiente de difusión y sedimentación, viscosidad intrínseca y dinámica rotacional), y propiedades conformacionales, como el radio de giro, curvas de intensidad de dispersión de rayos X o neutrones y la distribución de distancias. Además, esta metodología puede extenderse al estudio de sistemas agregados o multi-molécula, reconocimiento enzima-sustrato, etc...



Publicaciones resultantes de la Tesis

1. G. M. Pavlov, D. Amorós, C. Ott, I.I. Zaitseva, J. García de la Torre y U.S. Schubert, “Hydrodynamic analysis of well-defined flexible linear macromolecules of low molecular mass”. *Macromolecules*, 42, 7447-7455 (2009).
2. D. Amorós, A. Ortega, S. E. Harding y J. García de la Torre. “Multi-scale calculation and global-fit analysis of hydrodynamic properties of biological macromolecules: determination of the overall conformation of antibody IgG molecules”. *European Biophysics Journal*, 39, 361-370 (2010).
3. J. García de la Torre, D. Amorós y A. Ortega “Intrinsic viscosity of bead models for macromolecules and nanoparticles”. *European Biophysics Journal*, 39, 381-388 (2010)
4. J. García de la Torre, A. Ortega, D. Amorós, R. Rodríguez Schmidt y J. G. Hernández Cifre. “Methods and tools for the prediction of hydrodynamic coefficients and other solution properties of flexible macromolecules in solution”. *Macromolecular Bioscience*, 10. 721-730 (2010).
5. A. Ortega, D. Amorós y J. García de la Torre. “Global fit and structure optimization of flexible and rigid macromolecules and nanoparticles from analytical ultracentrifugation and other dilute solution properties”. *Methods*, 54 115-123. (2011).
6. D. Amorós, A. Ortega y J. García de la Torre. “Hydrodynamic Properties of Worm-like Macromolecules: Monte Carlo Simulation and Global Analysis of Experimental Data” *Macromolecules*, 44. 5788-5797 (2011).
7. A. Ortega, D. Amorós y J. García de la Torre. “Prediction of hydrodynamic and other solution properties of rigid proteins from atomic and residue-level models”. *Biophysical Journal*, 101. 892-898 (2011).
8. D. Amorós, A. Ortega y J. García de la Torre. “Prediction of hydrodynamic and other solution properties of disordered proteins”. Manuscrito en elaboración.



Referencias

- [1] A. Horta. *Macromoléculas*, volume 1. UNED, Madrid, 1982.
- [2] M. Rubinstein and R.H. Colby. *Polymer Physics*. Oxford University Press, New York, 2003.
- [3] P. Debye. Molecular weight determination by light scattering. *J. Phys. Colloid. Chem.*, 51:18, 1947.
- [4] A. Einstein. On the movement of small particles suspended in a stationary liquid, demanded by the molecular-kinetic theory of heat. *Ann. Physik*, 17:549–560, 1905. (Traducción al inglés en “Investigation on the Theory of the Brownian Movement”, Dover, New York, 1956).
- [5] S.F. Sun. *Physical Chemistry of Macromolecules (2nd Ed.)*. John Wiley and Sons, New Jersey, 2004.
- [6] R.F. Probstein. *Physicochemical Hydrodynamics. An Introduction (2nd Ed.)*. Wiley, New Jersey, 2003.
- [7] H. Brenner. Coupling between the translational and rotational Brownian motions of rigid particles of arbitrary shape. II. General considerations. *J. Colloid Interface Sci.*, 23:407–436, 1967.
- [8] J. García de la Torre and V.A. Bloomfield. Hydrodynamic properties of complex, rigid, biological macromolecules. Theory and applications. *Quart. Rev. Biophys.*, 14:81–139, 1981.
- [9] J. García de la Torre. Rotational diffusion coefficients. In S. Krause, editor, *Molecular Electro-Optics*, pages 75–103. Ed. Plenum Press, New York, 1981.

- [10] B. Carrasco and J. García de la Torre. Hydrodynamic properties of rigid particles. Comparison of different modelling and computational procedures. *Biophys. J.*, 76:3044–3057, 1999.
- [11] K.E. van Holde, W. Johnson, and P. Ho. *Principles of Physical Biochemistry, 2nd edition*. Prentice Hall, Upper Saddle River, N.J., 1998.
- [12] J. García de la Torre, M.L. Huertas, and B. Carrasco. Calculation of hydrodynamic properties of globular proteins from their atomic-level structures. *Biophys. J.*, 78:719–730, 2000.
- [13] J. García de la Torre, M.L. Huertas, and B. Carrasco. HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.*, 147:138–146, 2000.
- [14] P. Hagerman and B.H. Zimm. Monte Carlo approach to the analysis of the rotational diffusion of wormlike chains. *Biopolymers*, 20:1481–1502, 1981.
- [15] J. García de la Torre, S.E. Harding, and B. Carrasco. Calculation of NMR relaxation, covolume and scattering-related properties of bead models using the SOLPRO computer program. *Eur. Biophys. J.*, 28:119–132, 1999.
- [16] G. Lipari and A. Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.*, 104:4546–4559, 1982.
- [17] N. Tjandra, P. Wingfield, S. Stahl, and A. Bax. Anisotropic rotational diffusion of perdeuterated HIV protease from ^{15}N NMR relaxation measurements at two magnetic fields. *J. Biomol. NMR*, 8:273–284, 1996.
- [18] M. Buck, J. Boyd, C. Redfield, D.A. MacKenzie, D.J. Jeenes, D.B. Archer, and C.M. Dobson. Structural determinants of protein dynamics: analysis of ^{15}N NMR relaxation measurements for main-chain and side-chain nuclei of hen egg white lysozyme. *Biochemistry*, 34:4041–4055, 1995.
- [19] C. Tanford. *Physical Chemistry of Macromolecules*. J. Wiley and Sons, New York, 1961.

-
- [20] D. Freifelder, editor. *Physical Biochemistry. Applications to Biochemistry and Molecular Biology*. Freeman, New York, second edition, 1982.
- [21] A. Ortega and J. García de la Torre. Equivalent radii and ratios of radii from solution properties as indicators of macromolecular conformation, shape, and flexibility. *Biomacromolecules*, 8:2464–2475, 2007.
- [22] J. García de la Torre, S. Navarro, M.C. López Martínez, F.G. Díaz, and J.J. López Cascales. HYDRO: a computer software for the prediction of hydrodynamic properties of macromolecules. *Biophys. J.*, 67:530–531, 1994.
- [23] J. García de la Torre, G. Del Río Echenique, and A. Ortega. Improved calculation of rotational diffusion and intrinsic viscosity of bead models for macromolecules and nanoparticles. *J. Phys. Chem. B*, 111:955–961, 2007.
- [24] 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.*, 28:235–242, 2000.
- [25] J. García de la Torre and B. Carrasco. Hydrodynamic properties of rigid macromolecules composed of ellipsoidal and cylindrical subunits. *Biopolymers*, 63:163–167, 2002.
- [26] J. García de la Torre, H.E. Pérez Sánchez, A. Ortega, J.G. Hernández, M.X. Fernandes, F.G. Díaz, and M.C. López Martínez. Calculation of the solution properties of flexible macromolecules: methods and applications. *Eur. Biophys. J.*, 32:477–486, 2003.
- [27] J. E. Nordlander. A concise fortran program for molecular atomic coordinates. *J. Chem. Ed.*, 50 (11):743, 1973.
- [28] N. Metropolis, A.W. Metropolis, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculation by fast computing machines. *J. Chem. Phys.*, 21:1087–1093, 1953.
- [29] J. García de la Torre, A. Ortega, H.E. Pérez Sánchez, and J.G. Hernández Cifre. MULTIHIDRO and MONTEHYDRO: Conformational search and Monte Carlo calculation of solution properties of rigid and flexible macromolecular models. *Biophys. Chem.*, 116:121–128, 2005.

- [30] D.L. Ermak and J.A. McCammon. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.*, 69:1352–1360, 1978.
- [31] A. Iniesta and J. García de la Torre. A second-order algorithm for the simulation of the Brownian dynamics of macromolecular models. *J. Chem. Phys.*, 92:2015–2019, 1990.
- [32] P.J. Rossky, J.D. Doll, and H.L. Friedman. Brownian dynamics as a smart Monte Carlo method. *J. Chem. Phys.*, 69:4628–4634, 1978.
- [33] J. García de la Torre, H.E. Pérez Sánchez, A. Ortega, J.G. Hernández Cifre, M.X. Fernandes, F.G. Díaz Baños, and M.C. López Martínez. Calculation of the solution properties of flexible macromolecules: methods and applications. *Eur. Biophys. J.*, 32:477–486, 2003.
- [34] J. García de la Torre, A. Ortega, and H. E. Pérez. Solution properties of flexible macromolecules. theoretical and computacional approaches. In D. J. Scott, S.E. Harding, and A. Rowe, editors, *Analytical Ultracentrifugation: Techniques and methods*, pages 449–467. Royal Society of Chemistry, Oxford, 2005.
- [35] J. García de la Torre, J.G. Hernández Cifre, A. Ortega, R. Rodríguez Schmidt, M.X. Fernandes, H. E. Pérez Sánchez, and R. Pamies. SIMUFLEX : Algorithms and tools for simulation of the conformation and dynamics of flexible molecules and nanoparticles in solution. *J. Chem. Theor. Comput.*, 5:2606–2618, 2009.
- [36] M.A. Haney. A new differential viscometer - Part one. *American Laboratory*, 17:41–56, 1985.
- [37] J. Lee and A. Tripathi. Intrinsic viscosity of polymers and biopolymers measured by microchip. *Anal. Chem.*, 77:7137–7147, 2005.
- [38] E. Longman, S.E. Harding, and N. Marheineke. Identifying differences in solution conformations of two chimeric IgG3 antibodies through triple detection sec. *LCGC North America*, pages 86–90, 2006.
- [39] O.F. Solomon and I.Z. Ciuta. Détermination de la viscosité intrinsèque de solutions de polymères par une simple détermination de la viscosité. *J. Appl. Polym. Sci.*, 6:683–686, 1962.

-
- [40] R. Pamies, J.G. Hernández Cifre, M.C. López Martínez, and J. García de la Torre. Determination of intrinsic viscosities of macromolecules and nanoparticles. Comparison of single-point and dilution procedures. *Colloid Polym. Sci.*, 286:1223–1231, 2008.
- [41] J.G. Kirkwood and J. Riseman. The intrinsic viscosities and diffusion constants of flexible macromolecules in solution. *J. Chem. Phys.*, 16:565–573, 1948.
- [42] J. Riseman and J.G. Kirkwood. The intrinsic viscosity, translational and rotatory diffusion constants of rod-like macromolecules in solution. *J. Chem. Phys.*, 18:512–516, 1950.
- [43] H. Yamakawa. *Modern Theory of Polymer Solutions*. Harper and Row, New York, 1971.
- [44] J. García de la Torre and V.A. Bloomfield. Hydrodynamic properties of macromolecular complexes. I. Translation. *Biopolymers*, 16:1747–1763, 1977.
- [45] J. García de la Torre and V.A. Bloomfield. Hydrodynamic properties of macromolecular complexes. IV. Intrinsic viscosity theory with applications to once-broken rods and multisubunit proteins. *Biopolymers*, 17:1605–1627, 1978.
- [46] J. García de la Torre and B. Carrasco. Intrinsic viscosity and rotational diffusion of bead models for rigid particles. *Eur. Biophys. J.*, 27:549–557, 1998.
- [47] R. W. Wilson and V.A. Bloomfield. Hydrodynamic properties of macromolecular complexes. V. Improved calculation of rotational diffusion coefficient and intrinsic viscosity. *Biopolymers*, 18:1205–1211, 1979.
- [48] J.M. García Bernal and J. García de la Torre. Transport properties of oligomeric subunit structures. *Biopolymers*, 20:129–139, 1981.
- [49] M. Fixman. Variational bounds for polymer transport coefficients. *J. Chem. Phys.*, 78:1594–1599, 1983.
- [50] J.J. Freire and A. Rey. Lower bounds for the intrinsic viscosity of flexible polymers. comparison with Brownian dynamics simulation results for different types of chains. *Comput. Phys. Commun.*, 61:297–303, 1990.

- [51] G. Wei and B.E. Eichinger. On shape asymmetry of gaussian molecules. *J. Chem. Phys.*, 93:1430–1435, 1990.
- [52] A. Ortega and J. García de la Torre. Hydrodynamic properties of rodlike and disklike particles in dilute solution. *J. Chem. Phys.*, 119:9914–9919, 2003.
- [53] J. García de la Torre, M.C. López Martínez, M.M. Tirado, and J.J. Freire. Approximate methods for calculating hydrodynamic properties of macromolecules in dilute solution. Theory and application to rigid structures. *Macromolecules*, 16:1121–1127, 1983.
- [54] J. Riseman and J.G. Kirkwood. The statistical mechanical theory of irreversible processes in solutions of macromolecules. In *Rheology-Theory and Applications*, volume 1, chapter 13, page 495. 1956.
- [55] H. Yamakawa and G. Tanaka. Translational diffusion coefficients of rodlike polymers: Application of the modified oseen tensor. *J. Chem. Phys.*, 57:1537–1546, 1972.
- [56] J.G. Kirkwood and P.L. Auer. The visco-elastic properties of solutions of rod-like macromolecules. *J. Chem. Phys.*, 19:281–287, 1951.
- [57] H. Yamakawa. Viscoelastic properties of straight cylindrical macromolecules in dilute solution. *Macromolecules*, 8:339–342, 1975.
- [58] M.M. Tirado and J. García de la Torre. Translational friction coefficients of rigid, symmetric top macromolecules. *J. Chem. Phys.*, 71:2581–2587, 1979.
- [59] H. Yamakawa and J. Yamaki. Application of the Kirkwood theory of transport in polymer solutions to rigid assemblies of beads. *J. Chem. Phys.*, 58:2049–2055, 1973.
- [60] J. Ryckaert and A. Bellemans. Molecular dynamics of liquid n-butane near its boiling point. *Chem. Phys. Lett.*, 30:123–125, 1975.
- [61] G. Lamm and A. Szabo. Langevin modes of macromolecules. *J. Chem. Phys.*, 85:7334–7348, 1986.
- [62] R. G. Winkler, T. Matsuda, and D. Y. Yoon. Stochastic dynamics simulations of polymethylene melts confined between solid surfaces. *J. Chem. Phys.*, 98:729, 1992.

-
- [63] A. Rey, J.J. Freire, and J. García de la Torre. Brownian dynamics simulation of flexible polymer chains with excluded volume and hydrodynamic interaction. *Polymer*, 33:3477–3482, 1992.
- [64] H. Taketomi, Y. Ueda, and N. Gō. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res*, 7:445–459, 1975.
- [65] C. Clementi, H. Nyemeyer, and J. Onuchic. Topological and energetic factors: What determines the structural details of the transition state assemble and en-route intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.*, 298:937–53, 2000.
- [66] V. Tozzini. Mapping all-atom models onto one-bead coarse-grained models: General properties and applications to a minimal polypeptide model. *J. Chem. Theory Comput.*, 2:667–673, 2006.
- [67] J. García de la Torre, O. Llorca, J.L. Carrascosa, and J.M. Valpuesta. HYDROMIC: Prediction of hydrodynamic properties of rigid macromolecular structures obtained from electron microscopy. *Eur. Biophys. J.*, 30:457–462, 2001.
- [68] A. Ortega and J. García de la Torre. Efficient, accurate calculation of rotational diffusion and NMR relaxation of globular proteins from atomic-level structures and approximate hydrodynamic calculations. *J. Am. Chem. Soc.*, 127:12764–12765, 2005.
- [69] P. Zipper and H. Durchschlag. Modeling of protein solution structures. *J. Appl. Cryst.*, 36:509–514, 2003.
- [70] N. Rai, M. Nollman, B. Spotorno, G. Tassara, O. Byron, and M. Rocco. SOMO (Solution MOdeller) differences between X-ray and NMR-derived bead models suggest a role for side chain flexibility in protein dynamics. *Structure*, 13:722–734, 2005.
- [71] P. Bernadó, J. García de la Torre, and M. Pons. Interpretation of ^{15}N NMR relaxation data of globular proteins using hydrodynamic calculations with HYDRONMR. *J. Biomol. NMR*, 23:139–150, 2002.

- [72] J. García de la Torre. Building hydrodynamic bead-shell models for rigid particles of arbitrary shape. *Biophys. Chem.*, 94:265–274, 2001.
- [73] M.X. Fernandes, P. Bernadó, M. Pons, and J. García de la Torre. An analytical solution to the problem of the orientation of rigid particles by planar obstacles. *J. Am. Chem. Soc.*, 123:12037–12047, 2001.
- [74] Álvaro Ortega Retuerta. *Metodologías computacionales para propiedades en disolución de macromoléculas rígidas y flexibles*. PhD thesis, Universidad de Murcia, 2005.
- [75] A. Ortega and J. García de la Torre. Equivalent radii and ratios of radii from solution properties as indicators of macromolecular shape, conformation and flexibility. *Biomacromolecules*, 8:2464–2475, 2007.
- [76] B. Carrasco, J. García de la Torre, O. Byron, D. King, C. Walters, S. Jones, and S.E. Harding. Novel size-independent modeling of the dilute solution conformation of the immunoglobulin IgG Fab’ domain using SOLPRO and ELLIPS. *Biophys. J.*, 77:2902–2910, 1999.
- [77] B. Carrasco, J. García de la Torre, K.G. Davis, S. Jones, D. Athwal, C. Walters, D.R. Burton, and S.E. Harding. Crystallohydrodynamics for solving the hydration problem for multi-domain proteins: open physiological conformations of human IgG. *Biophys. Chem.*, 93:181–196, 2001.
- [78] E. Longman, K. Kreusel, S.B. Tendler, I. Fiebig, K. King, J. Adiar, P. O’Shea, A. Ortega, and J. García de la Torre. Estimating domain orientation of two human antibody IgG4 chimeras by crystallohydrodynamics. *Eur. Biophys. J.*, 32:502–510, 2003.
- [79] S.E. Harding, E. Longman, B. Carrasco, A. Ortega, and J. García de la Torre. Studying antibody conformations by ultracentrifugation and hydrodynamic modeling. In *Methods in Molecular Biology*, volume 248, chapter 5, pages 93–113. Humana Press Inc., Totowa, N.J., 2003.
- [80] S.E. Harding, E. Longman, A. Ortega, K. Kreusel, S.B. Tendler, K.K. King, and J. García de la Torre. Use of the sedimentation coefficient for modelling antibodies.

- Refinements to the crystallohydrodynamics approach. *Progr. Colloid and Polym. Sci.*, 127:113–119, 2004.
- [81] Y. Lu, E. Longman, K.G. Davis, A. Ortega, G.G. Grossmann, T.E. Michelsenand, and J. García de la Torre. Crystallohydrodynamics of protein assemblies: combining sedimentation, viscometry and X-ray scattering. *Biophys. J.*, 91:1688–1697, 2006.
- [82] Y. Lu, S. Harding, T. Michaelsen, E. Longman, K.G. Davis, A. Ortega, G. Grossman, I. Sandlie, and J. García de la Torre. Solution conformation of wild type and mutant IgG3 and IgG4 immunoglobulins using crystallohydrodynamics: posible implications for complement activation. *Biophys. J.*, 93:3733–3744, 2007.
- [83] R.L. Brady, D.J. Edwards, R.E. Hubbard, J.S. Jiang, G. Lange, S.M. Roberts, R.J. Todd, J.R. Adair, J.S. Emtage, D.J. King, and D.C. Low. Crystal structure of a chimeric Fab' fragment of an antibody binding tumour cells. *J. Mol. Biol.*, 227:253–264, 1992.
- [84] J. Deisenhofer. Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureus* at 2.9 Å and 2.8 Å resolution. *Biochemistry*, 20:2361–2370, 1981.
- [85] O. Byron. *Solution studies on the conformation and assembly of the monoclonal antibody B72.3*. PhD thesis, University of Nottingham, 1992.
- [86] A. Sólvez, A. Iniesta, and J. García de la Torre. Radius of gyration of multisubunits macromolecules. Theory and application to myosin heads, myosin rod and whole myosin. *Int. J. Biol. Macromolecules*, 9:39–43, 1988.
- [87] O. Kratky and G. Porod. Rontgenuntersushung geloster fagenmolekule. *Rec. Trav. Chim.*, 68:1106–1122, 1949.
- [88] A. D. McNaught and A. Wilkinson. *Compendium of Chemical Terminology. The Gold Book, Second Edition*. Blackwell Science, London, 1997.
- [89] V.A. Bloomfield, D.M. Crothers, and I. Tinoco. *Physical Chemistry of Nucleic Acids*. Harper and Row, New York, 1974.
- [90] V.A. Bloomfield, D.M. Crothers, and I. Tinoco Jr. *Nucleic Acids: Structures, Properties and Functions*. University Science Books, Sausalito, California, 2000.

- [91] T. Sho, T. Sato, and T. Norisuye. Viscosity behavior and persistence length of sodium xanthan. *Biophys. Chem.*, 25:307–313, 1986.
- [92] S.E. Harding, G. Berth, A. Ball, J.R. Mitchell, and J. García de la Torre. The molecular weight distribution and conformation of citrus pectins in solution studied by hydrodynamics. *Carbohydr. Polym.*, 16:1–15, 1991.
- [93] H. Yamakawa and M. Fujii. Translational friction coefficient of wormlike chains. *J. Chem. Phys.*, 6:408–414, 1973.
- [94] H. Yamakawa and M. Fujii. Intrinsic viscosity of wormlike chains. Determination of the shift factor. *Macromolecules*, 7:128–135, 1974.
- [95] H. Yamakawa and M. Fujii. Light scattering from wormlike chains. Determination of the shift factor. *Macromolecules*, 7:649–654, 1974.
- [96] J. García de la Torre, M.C. López Martínez, M.M. Tirado, and J.J. Freire. Monte Carlo study of hydrodynamic properties of flexible linear chains. Analysis of several approximate methods. *Macromolecules*, 17:2715–2722, 1984.
- [97] B.H. Zimm. Chain molecule hydrodynamics by the Monte-Carlo method and the validity of the Kirkwood-Riseman approximation. *Macromolecules*, 13:592–602, 1980.
- [98] J. García de la Torre, A. Jiménez, and J.J. Freire. Monte Carlo calculation of hydrodynamic properties of freely jointed, freely rotating and real polymethylene chains. *Macromolecules*, 15:148–154, 1982.
- [99] J.M. García Bernal, M.M. Tirado, J.J. Freire, and J. García de la Torre. Monte Carlo calculation of hydrodynamic properties of cyclic polymers in ideal solution. *Macromolecules*, 23:3357–3362, 1990.
- [100] J.M. García Bernal, M.M. Tirado, and J. García de la Torre. Monte Carlo calculation of hydrodynamic properties of linear and cyclic chains in good solvents. *Macromolecules*, 24:593–598, 1991.
- [101] J.J. García Molina, M. C. López Martínez, and J. García de la Torre. Computer simulation of hydrodynamic properties of semiflexible macromolecules. Randomly broken chains, wormlike chains, and analysis of properties of DNA. *Biopolymers*, 29:883–900, 1990.

-
- [102] H. Benoit and P.M. Doty. Light scattering from non-Gaussian chains. *J. Phys. Chem.*, 57:958–963, 1953.
- [103] B. Carrasco and J. García de la Torre. Improved hydrodynamic interaction in macromolecular bead models. *J. Chem. Phys.*, 111:4817–4826, 1999.
- [104] J.A. Schellman. Flexibility of DNA. *Biopolymers*, 13:217–226, 1974.
- [105] P.J. Flory. *Statistical Mechanics of Chain Molecules*. John Wiley, New York, 1st edition, 1969.
- [106] P.J. Flory. *Principles of Polymer Chemistry*. Cornell University Press, Ithaca, NY, 1953.
- [107] P.J. Flory and T. G. Fox Jr. Treatment of intrinsic viscosities. *J. Am. Chem. Soc.*, 73:1904–1908, 1951.
- [108] A. M. Jamieson and K. Venkataswamy. Static and dynamic radii of polystyrene in ethylbenzene and tetrahydrofuran. *Polym. Bull.*, 12:275–282, 1984.
- [109] K. Venkataswamy, A.M. Jamieson, and R.G. Petschek. Static and dynamic properties of polystyrene in good solvents: ethylbenzene and tetrahydrofuran. *Macromolecules*, 19:124–133, 1986.
- [110] R. P. Brent. *Algorithms for minimization without derivatives*. Prentice-Hall, Englewood Cliffs, 1986.
- [111] Y. Lu, B. Weers, and N. C. Stellwagen. DNA persistence length revisited. *Biopolymers*, 61:261–275, 2002.
- [112] J. G. Elias and D. Eden. Transient electric birefringence study of the persistence length and electrical polarizability of restriction fragments of DNA. *Macromolecules*, 14:410–419, 1981.
- [113] S. Diekmann, W. Hillen, B. Morgeneyer, R.D. Wells, and D. Pörschke. Orientation relaxation of DNA restriction fragments and the internal mobility of the double helix. *Biophys. Chem.*, 61:261–275, 2002.
- [114] S.S. Sorlie and R. Pecora. A dynamic light scattering study of a 2311 base pair DNA restriction fragment. *Macromolecules*, 21:1437–1449, 1988.

- [115] S.A. Allison, S.S. Soarlie, and R. Pecora. Brownian dynamics simulation of wormlike chains. Dynamic light scattering from a 2311 base pairs fragment. *Macromolecules*, 23:1110–1114, 1990.
- [116] D. Jolly and H. Eisenberg. Photon correlation spectroscopy, total intensity light-scattering with laser radiation, and hydrodynamic studies of a well fractionated DNA sample. *Biopolymers*, 15:61–95, 1976.
- [117] Z. Kam, N. Borochoy, and H. Eisenberg. Dependence of laser light scattering of DNA on NaCl concentration. *Biopolymers*, 20:2671–2690, 1981.
- [118] S.S. Sorlie and R. Pecora. A dynamic light scattering study of four DNA restriction fragments. *Macromolecules*, 23:487–497, 1990.
- [119] K. Soda and A. Wada. Dynamic light-scattering-studies on thermal motions of native DNAs in solution. *Biophys. Chem.*, 20:185–200, 1984.
- [120] J. Wilcoxon and J. M. Schurr. Dynamic light scattering from thin rigid rods: Anisotropy of translational diffusion of tobacco mosaic virus. *Biopolymers*, 22:849–867, 1983.
- [121] T. Nicolai and M. Mandel. Dynamic light scattering by aqueous solutions of low molar mass DNA fragments in the presence of NaCl. *Macromolecules*, 22:2348–2356, 1989.
- [122] W. Eimer, J.R. Williamson, S.G. Boxer, and R. Pecora. Characterization of the overall and internal dynamics of short oligonucleotides by depolarized dynamic light scattering and NMR relaxation measurements. *Biochemistry*, 29:799–811, 1990.
- [123] W. Eimer and R. Pecora. Rotational and translational diffusion of short rodlike molecules in solution: oligonucleotides. *J. Chem. Phys.*, 94:2324–2329, 1991.
- [124] G.F. Bonifacio, T. Brown, G.L. Conn, and A.N. Lane. Comparison of the electrophoretic and hydrodynamic properties of DNA and RNA oligonucleotide duplexes. *Biophys. J.*, 73:1532–1538, 1997.
- [125] N.C. Stellwagen, S. Magnusdottir, C. Gelfi, and P.G. Righetti. Measuring the translational diffusion coefficients of small DNA molecules by capillary electrophoresis. *Biopolymers*, 58:390–397, 2001.

-
- [126] J. Seils and T. Dorfmueller. Internal dynamics of linear and superhelical DNA as studied by photn-correlation spectroscopy. *Biopolymers*, 31:813–825, 1991.
- [127] A.I. Krasna. Changes in the light-scattering properties of DNA on denaturation. *J. Colloid Interface Sci.*, 39:632–646, 1972.
- [128] N. Borochoy, H. Eisenberg, and Z. Kam. Dependence of DNA conformation on the concentration of salt. (Communications to the Editor). *Biopolymers*, 20:231–235, 1981.
- [129] J.A. Harpst. Analysis of low angle light scattering results from T7 DNA. *Biophys. Chem.*, 11:295–302, 1980.
- [130] J.E. Godfrey. The flexibility of low molecular weight double-stranded DNA as a function of length. (I). *Biophys. Chem.*, 5:285–299, 1976.
- [131] A.I. Krasna, J. R. Dawson, and J. A. Harpst. Characterization of acid-denatured DNA by low-angle light scattering. *Biopolymers*, 9:1017–1028, 1970.
- [132] C.W. Schmid, F.P. Rineheart, and J.B. Hearst. Statistical length of DNA from light scattering. *Biopolymers*, 10:883–893, 1971.
- [133] M.A. Llorente and A. Horta. Extrapolation of light-scattering data to determine the molecular weight of T7 DNA. *Biopolymers*, 15:2287–2290, 1976.
- [134] H. I. Levine, R. J. Fiel, and F. W. Billmeyer Jr. Very low-angle light scattering. A characterization method for high-molecular-weight DNA. *Biopolymers*, 15:1267–1281, 1976.
- [135] J.A. Harpst, A.I. Krasna, and B.H. Zimm. Molecular weight of T7 and calf thymus DNA by low-angle light scattering. *Biopolymers*, 6:595–603, 1968.
- [136] M. Mandel and J. Schouten. Ionic strength dependence of the average dimension of low molecular weight DNA. *Macromolecules*, 13:1247–1251, 1980.
- [137] R.G. Kirste. Radius of gyration of stiff chain molecules as a function of the chain length and the interactions with the solvent. *Discuss. Faraday Soc.*, 49:51, 1970.
- [138] P. Doty. The properties of sonic fragments of deoxyribose nucleic acid. *Proc. Natl. Acad. Sci. U.S.A.*, 44:432–438, 1958.

REFERENCIAS

- [139] G. Cohen and H. Eisenberg. Conformation studies on the sodium and cesium salts of calf thymus deoxyribonucleic acid (DNA). *Biopolymers*, 4:429–440, 1966.
- [140] J.A. Harpst and J.R. Dawson. Low angle light scattering studies on whole, half and quarter molecules of T2 bacteriophage DNA. *Biophys. J.*, 55:1237–1249, 1989.
- [141] R.T. Kovacic and K.E. Van Holde. Sedimentation of homogeneous double stranded DNA molecules. *Biochemistry*, 16:1490–1498, 1977.
- [142] H.B. Gray Jr. and J.E. Hearst. Flexibility of native DNA from the sedimentation behavior as a function of molecular weight and temperature. *J. Mol. Biol.*, 35:111–129, 1968.
- [143] A.H. Rosemberg and F.W. Studier. Intrinsic viscosity of native and single-stranded T7 DNA and its relationship to sedimentation coefficient. *Biopolymers*, 7:765–774, 1969.
- [144] M.T. Record, C.P. Woodbury, and R. Inman. Characterization of rodlike DNA fragments. *Biopolymers*, 14:393–408, 1975.
- [145] J.E. Hearst, C.W. Schmid, and F.P. Rineheart. Molecular weights of homogeneous samples of desoxyribonucleic acid determined from hydrodynamic theories for the wormlike coil. *Macromolecules*, 1:491–494, 1968.
- [146] D. Crothers and B. Zimm. Viscosity and sedimentation of DNA from bacteriophages T2 and T7 and relation to molecular weight. *J. Mol. Biol.*, 12:525–536, 1965.
- [147] I. Noda. Sedimentation coefficient and intrinsic viscosity of semiflexible chains. *Biopolymers*, 13:591–599, 1974.
- [148] R.E. Harrington. Optico-hydrodynamic properties of high molecular weight DNA from steady-state flow birefringence and viscosity at extremely low velocity gradients. *Biopolymers*, 9:159–193, 1970.
- [149] C.F. Reeg and R. E. Harrington. Steady-state opticohydrodynamic properties of DNA molecular weight dependence and the internal viscosity problem. *Biopolymers*, 21:1315–1332, 1982.

-
- [150] P.D. Ross and R.L. Scruggs. Viscosity study of DNA. II. The effect of simple salt concentration on the viscosity of high molecular weight DNA and application of viscometry to the study of DNA isolated from T4 and T5 bacteriophage mutants. *Biopolymers*, 6:1005–1018, 1968.
- [151] J.E. Godfrey and H. Eisenberg. The flexibility of low molecular weight double-stranded DNA as a function of length. (II). *Biophys. Chem.*, 5:301–318, 1976.
- [152] B.C. Bowen and B.H. Zimm. Molecular weight of T2 NaDNA from viscoelasticity. *Biophys. Chem.*, 7:235–252, 1978.
- [153] M. Tanigawa, N. Mukaiyama, S. Shimokubo, K. Wakabayashi, Y. Fujita, K. Fukudome, and K. Yamaoka. Molecular weight distribution and intrinsic viscosity of sonicated and successively fractionated double-stranded deoxyribonucleic acid and polyribonucleotides. *Polym. J.*, 26:291–302, 1994.
- [154] G.M. Pavlov, D. Amorós, C. Ott, I.I. Zaitseva, J. García de la Torre, and U.S. Schubert. Hydrodynamic analysis of well-defined flexible linear macromolecules of low molecular mass. *Macromolecules*, 42:7447–7455, 2009.
- [155] J.E. Hearst. Rotatory diffusion constants of stiff-chain macromolecules. *J. Chem. Phys.*, 38:1062–1065, 1962.
- [156] K. Kashiwagi, T. Norisuye, and H. Fujita. Triple helix of schizophyllum commune polysaccharide in dilute solution. 4. light scattering and viscosity in dilute aqueous solution. *Macromolecules*, 14:1220–1125, 1981.
- [157] T. Yanaki, T. Norisuye, and H. Fujita. Triple helix of schizophyllum commune polysaccharide in dilute solution. 3 hydrodynamic properties in water. *Macromolecules*, 13:1462–1466, 1980.
- [158] H. Murakami, T. Norisuye, and H. Fujita. Dimensional and hydrodynamic properties of poly(hexyl isocyanate) in hexane. *Macromolecules*, 13:345–352, 1980.
- [159] D. N. Rubingh and H. Yu. Characterizations of stiff chain macromolecules. poly(n-hexyl isocyanate) in n-hexane. *Macromolecules*, 9:681–685, 1976.

- [160] F.I Abe, Y. Einaga, and H. Yamakawa. Excluded-volume effects on the intrinsic viscosity of oligomers and polymers of styrene and isobutylene. *Macromolecules*, 26:1891–1897, 1993.
- [161] M. Osa, F. Abe, T. Yoshizaki, Y. Einaga, and H. Yamakawa. Excluded-volume effects on the hydrodynamic radius of oligo and polyisobutylenes in dilute solution. *Macromolecules*, 29:2302–2308, 1996.
- [162] M. Yamada, M. Osa, T. Yoshizaki, and H. Yamakawa. Excluded-volume effects on the mean-square radius of gyration of oligo and polyisobutylenes in dilute solution. *Macromolecules*, 30:7166–7170, 1997.
- [163] L. J. Fetters, N. Hadjichristidis, J. S. Lindner, J. W. Mays, and W. W. Wilson. Transport properties of polyisobutylene in dilute solution. *Macromolecules*, 24:3127–3135, 1991.
- [164] M. L. Mansfield and J. F. Douglas. Transport properties of wormlike chains with applications to double helical DNA and carbon nanotubes. *Macromolecules*, 41:5412–5421, 2008.
- [165] A. K. Dunker, J. D. Lawson, C. J. Brown, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, and Z. Obradovic. Intrinsically disordered protein. *J. Mol. Graph. Model.*, 19:26–59, 2001.
- [166] P. Tompa. Intrinsically unstructured proteins. *Trends Biochem. Sci.*, 27:527–533, 2002.
- [167] E. Fischer. Einfluss der configuration auf die wirkung der enzyme. *Ber. Dt. Chem. Ges.*, 27:2985–2993, 1894.
- [168] J. García de la Torre. Hydration from hydrodynamics. General considerations and applications of bead modelling to globular proteins. *Biophys. Chem.*, 93:150–159, 2001.
- [169] D.K. Hahn and S.R. Aragon. Intrinsic viscosity of proteins and platonic solids by boundary element methods. *J. Chem. Theory Comput.*, 2:1416–1428, 2006.

-
- [170] E. Brookes, B. Demeler, and M. Rocco. Developments in the US-SOMO bead modeling suite: New features in the direct residue-to-bead method, improved grid routines, and influence of accessible surface area screening. *Macromol. Biosci.*, 10:746–753, 2010.
- [171] D. K. Clare, P. J. Bakkes, H. Van Heerikhuizen, S. M. Van Der Vies, and H. R. Saibil. An expanded protein folding cage in the GroEL-gp31 complex. *J. Mol. Biol.*, 358:905–911, 2006.
- [172] J. Behlke, O. Ristau, and H. J. Schönfeld. Nucleotide-dependent complex formation between the *Escherichia coli* chaperonins GroEL and GroES studied under equilibrium conditions. *Biochemistry*, 36:5149–5156, 1997.
- [173] I. Tomonao, K. Takahashi, K. Maki, S. Enoki, K. Kamagata, A. Kadooka, M. Arai, and K. Kuwajima. Asymmetry of the GroEL-GroES complex under physiological conditions as revealed by small-angle X-ray scattering. *Biophys. J.*, 94:1392–1402, 2008.
- [174] A. Balasubramanian and K. Ponnuraj. Crystal structure of the first plant urease from jack bean: 83 years of journey from its first crystal to molecular structure. *J. Mol. Biol.*, 400:274–283, 2010.
- [175] W. N. Fishbein, K. Nagarajan, and W. Scurzi. Urease catalysis and structure. VI. Correlation of sedimentation coefficients and electrophoretic mobilities for the polymeric urease isozymes. *J. Biol. Chem.*, 215:5985–5992, 1970.
- [176] M.L. Huertas, S. Navarro, M.C. López Martínez, and J. García de la Torre. Simulation of the conformation and dynamics of a double-helicoidal model for DNA. *Biophys. J.*, 73:3142–3153, 1997.
- [177] V. B. Patel, C. C. Cunningham, and R. R. Hantgan. Physiochemical properties of rat liver mitochondrial ribosomes. *J. Biol. Chem.*, 276:6739–6746, 2001.
- [178] W. E. Hill, G. P. Rossetini, and K. E. Van Holde. Physical studies of ribosomes from *Escherichia coli*. *J. Mol. Biol.*, 44:263–277, 1969.
- [179] S. J. Perkins, A. S. Nealis, B. J. Sutton, and A. Feinstein. Solution structure of human and mouse immunoglobulin M by synchrotron X-ray scattering and molecular

- graphics modelling. a possible mechanism for complement activation. *J. Mol. Biol.*, 221:1345–1366, 1991.
- [180] J. K. Armstrong, R. B. Wenby, H. J. Meiselman, and T. C. Fisher. The hydrodynamic radii of macromolecules and their effect on red blood cell aggregation. *Biophys. J.*, 87:4259–4270, 2004.
- [181] J. B. Sumner, N. Gralen, and I. B. Eriksson-Quensel. The molecular weight of urease. *J. Biol. Chem.*, 125:37–44, 1932.
- [182] W. N. Fishbein, K. Nagarajan, and W. Scurzi. Urease catalysis and structure. IX. The half unit and hemipolymers of jack bean urease. *J. Biol. Chem.*, 248:7870–7877, 1973.
- [183] R. E. Schrohenloher and J. C. Bennett. Degradation of human IgM by pepsin: Characterization of a high molecular weight fragment. *J. Immunol.*, 107:870–880, 1971.
- [184] P. Wilhelm, I. Pilz, W. Palm, and K. Bauer. Small-angle X-ray studies of a human immunoglobulin M. *Eur. J. Biochem.*, 84:457–463, 1978.
- [185] A. R. Scafati, M. R. Stornaiuolo, and P. Novaro. Physicochemical and light scattering studies on ribosome particles. *Biophys. J.*, 11:370–384, 1971.
- [186] D. E. Koppel. Study of *Escherichia coli* ribosomes by intensity fluctuation spectroscopy of scattered laser light. *Biochemistry*, 13:2712–2719, 1974.
- [187] G. J. Kleywegt. Validation of protein models from C α coordinates alone. *J. Mol. Biol.*, 273:371–376, 1997.
- [188] R. W. W. Hooft, C. Sander, and G. Vriend. Verification of protein structures: Side-chain planarity. *Journal of Applied Crystallography*, 29:714–716, 1996.
- [189] J. E. Kohn, I. S. Millett, J. Jacob, B. Zagrovic, T. M. Dillon, N. Cingel, R. S. Dothager, S. Seifert, P. Thiyagarajan, T. R. Sosnick, M. Z. Hasan, V. S. Pande I. Ruczinski, S. Doniach, and K. W. Plaxco. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 101:12491–12496, 2004.

-
- [190] P. Bernadó and M. Blackledge. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and Small Angle Scattering. *Biophys. J.*, 97:2839–2845, 2009.
- [191] C. Tanford, K. Kawahara, and S. Lapanje. Proteins as random coils. I. Intrinsic viscosities and sedimentation coefficients in concentrated guanidine hydrochloride. *J. Am. Chem. Soc.*, 89:729–736, 1967.
- [192] C. Tanford. Protein denaturation. *Adv. Protein. Chem.*, 23:122–182, 1968.
- [193] S. Lapange. Random coil behavior of proteins in concentrated urea solutions. *Croat. Chem. Acta*, 41:115, 1969.
- [194] F. Ahmad and A. Salahuddin. Influence of temperature on the intrinsic viscosities of proteins in random coil conformation. *Biochemistry*, 13:245–249, 1974.
- [195] H. Zhou. Dimensions of denatured protein chains from hydrodynamic data. *J. Phys. Chem. B*, 106:5769–5775, 2004.
- [196] D. K. Wilkins, S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones, and L. J. Smith. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry*, 38:16424–16431, 1999.
- [197] K. Gast, G. Damashun, H. Damashun, R. Misselwitz, D. Zirwer, and V. E. V. E. Bychkova. *Laser Light Scattering in Biochemistry (S. Harding and D. B. Sattelle, Eds.)*. Royal Society of Chemistry, 1991.
- [198] Y. Izumi, Y. Miyake, K. Kuwajima, S. Sugai, K. Inoue, M. Iizumi, and S. Katano. Folding-unfolding of α -lactalbumin. *Physica B+C*, 120:444–448, 1983.
- [199] V. E. Bychkova, S. F. Bartoshevich, and S. L. Klenin. A comparative investigation of diffusion coefficients of α -lactalbumins and lysozyme by means of polarization interferometer. *Biofizika (Moscow)*, 35:242, 1990.
- [200] O. Tcherkasskaya and V. N. Uversky. Denatured collapsed states in protein folding: Example of apomyoglobin. *Proteins: Struct. Funct. Genet.*, 44:244–254, 2001.
- [201] V. N. Uversky. What does it mean to be natively unfolded? *Eur. J. Biochem.*, 269:2–12, 2002.

- [202] G. C. Berry. The hydrodynamic and conformational properties of denatured proteins in dilute solutions. *Protein Sci.*, 19:94–98, 2010.
- [203] V. Sobolev, A. Sorokine, J. Prilusky, E. E. Abola, and M. Edelman. Automated analysis of interatomic contacts in proteins. *Bioinformatics*, 15:327–332, 1999.
- [204] T. Frembgen-Kesner and A. H. Elcock. Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins. *J. Chem. Theory Comput.*, 5(2):242–256, 2009.
- [205] N. Eswar, M. A. Marti-Renom, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, and A. Sali. Comparative protein structure modeling with MODELLER. *Curr. Protoc. Bioinform.*, 15:5.6.1–5.6.30, 2006.
- [206] T. Frembgen-Kesner and A.H. Elcock. Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins. *J. Chem. Theory Comput.*, 5:242–256, 2009.
- [207] V. Tozzini, J. Trylska, C. Chang, and J. A. McCammon. Flap opening dynamics in HIV-1 protease explored with a coarse-grained model. *Journal of Structural Biology*, 157:606–615, 2007.
- [208] E. Caballero-Manrique, J. K. Bray, W. A. Deutschman, F. W. Dahlquist, and M. G. Guenza. A theory of protein dynamics to predict NMR relaxation. *Biophys. J.*, 93:4128–4140, 2007.
- [209] D. I. Svergun, C. Barberato C., and M. H. J. Koch. CRY SOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.*, 28:768–773, 1995.
- [210] P. Rotkiewicz and J. Skolnick. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J. Comput. Chem.*, 29:1460–1465, 2008.
- [211] M. Hammel, H-P. Fierobe, M. Czjzek, V. Kurkal, J. C. Smith, E. A. Bayer, S. Finet, and V. Reveceur-Bréchet. Structural basis of cellulosome efficiency explored by Small Angle X-ray Scattering. *J. Biol. Chem.*, 280:38562–38568, 2005.

-
- [212] L. Blanchard, N. Tarbouriech, M. Blackledge, P. Timmins, W. P. Burmeister, R. W. H. Ruigrok, and D. Marion. Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology*, 319:201–211, 2004.
- [213] P. Bernadó, L. Blanchard, P. Timmins, D. Marion, R. W. H. Ruigrok, and M. Blackledge. A structural model for unfolded proteins from residual dipolar couplings and Small-Angle X-ray Scattering. *Proc. Natl. Acad. Sci. U.S.A.*, 102:17002–17007, 2005.
- [214] K. Houben, L. Blanchard, M. Blackledge, and D. Marion. Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase cofactor P. *Biophys. J.*, 93:2830–2844, 2007.
- [215] I. Bertini, V. Calderone, M. Fragai, R. Jaiswal, C. Luchinat, M. Melikian, E. Mylonas, and D. I. Svergun. Evidence of reciprocal reorientation of the catalytic and hemopexin-like domains of full-length MMP-12. *J. Am. Chem. Soc.*, 130:7011–7021, 2008.
- [216] J. A. Marquez, C. I. Smith, M. V. Petoukhov, P. Lo Surdo, P. T. Mattsson, M. Knekt, A. Westlund, K. Scheffzek, M. Saraste, and D. I. Svergun. Conformation of full-length bruton tyrosine kinase (btk) from synchrotron X-ray solution scattering. *EMBO J.*, 22:4616–4624, 2003.
- [217] P. Bernadó, E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun. Structural characterization of flexible proteins using Small-Angle X-ray Scattering. *J. Am. Chem. Soc.*, 129:5656–5664, 2007.