



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE QUÍMICA

Estudio computacional del heterotetrámero de tubulinas y su interacción con estabilizadores microtubulares, con aplicación en el tratamiento de la enfermedad de Alzheimer

TESIS

QUE PARA OBTENER EL TÍTULO DE

QUÍMICO FARMACÉUTICO BIÓLOGO

PRESENTA

GUILLERMO DAVID GOODE ROMERO



MÉXICO, CDMX 2017



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO:

PRESIDENTE:	JOSÉ PEDRAZA CHAVERRI
VOCAL:	LAURA DOMÍNGUEZ DUEÑAS
SECRETARIO:	PEDRO JOSUÉ TREJO SOTO
1er. SUPLENTE:	JOSÉ LUIS MEDINA FRANCO
2° SUPLENTE:	FRANCISCA MORAYNA GUTIÉRREZ LUNA

SITIO DONDE SE DESARROLLÓ EL TEMA:

CUBÍCULO 212, EDIFICIO F, DEPARTAMENTO DE FISIQUÍMICA, FACULTAD DE QUÍMICA,
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

ASESOR DEL TEMA:

DRA. LAURA DOMÍNGUEZ DUEÑAS

SUPERVISOR TÉCNICO:

M. EN C. RODRIGO AGUAYO ORTIZ

SUSTENTANTE:

GUILLERMO DAVID GOODE ROMERO

AGRADECIMIENTOS

AGRADEZCO A LA DIRECCIÓN GENERAL DE CÓMPUTO Y DE TECNOLOGÍAS DE INFORMACIÓN, EN EL PROYECTO LANCAD-UNAM-DGTIC-306 POR EL ACCESO A LA SUPERCOMPUTADORA HP CLUSTER PLATFORM 3000SL (MIZTLI).

Y AL PROGRAMA DE APOYO A PROYECTOS DE INVESTIGACIÓN E INNOVACIÓN TECNOLÓGICA (PAPIIT) CON CLAVE IA204716.

ÍNDICE	i
SÍMBOLOS, MAGNITUDES, ABREVIATURAS Y ACRÓNIMOS	viii
I. RESUMEN E INTRODUCCIÓN	1
II. ANTECEDENTES	4
II.1. Generalidades	4
II.2. Etiología	4
II.2.1. Placas seniles	5
II.2.2. Ovillos neurofibrilares	6
II.3. Microtúbulos	7
II.3.1. Desestabilizadores microtubulares	8
II.3.2. Estabilizadores microtubulares	8
II.4. Barrera hematoencefálica	11
III. JUSTIFICACIÓN	12
IV. OBJETIVOS	12
IV.1. Objetivo general	12
IV.2. Objetivos particulares	12
V. METODOLOGÍA GENERAL	13
V.1. Primera parte. Selección de compuestos	14
V.1.1. Similitud molecular	14
V.1.2. Relaciones cuantitativas estructura-propiedad (QSPR)	14
V.1.3. Predicción del transporte a través de la BHE	15
V.2. Segunda parte. Acoplamiento y simulaciones de dinámica molecular	15
V.2.1. Alineamiento de secuencias	15
V.2.2. Modelado por homología	15
V.2.3. Construcción del sistema multimérico	15
V.2.3.1. Construcción del dímero $\alpha\beta$	15
V.2.3.2. Construcción del heterotetrámero $\alpha_2\beta_2$	15
V.2.4. Simulación de dinámica molecular	15
V.2.4.1. Parametrización de ligandos	16
V.2.4.2. Preparación del sistema	16
V.2.4.3. Minimización energética	16
V.2.4.4. Colectivo canónico (NVT)	16
V.2.4.5. Colectivo isobárico isotérmico (NPT)	16
V.2.4.6. Dinámica molecular (MD)	16

V.2.4.7 Análisis de MD del heterodímero y los heterotetrámeros	16
V.2.4.8. Agrupamiento o <i>clustering</i>	16
V.2.5. Acoplamiento molecular	17
V.2.6. Simulación MD de complejos proteína-ligando	18
VI. RESULTADOS	19
VI.1. Primera parte. Selección de compuestos	19
VI.1.1. Cribado virtual	19
VI.1.2. Relaciones cuantitativas estructura-actividad	19
VI.2. Segunda parte. Acoplamiento y simulaciones de dinámica molecular	22
VI.2.1. Alineamiento de secuencias	22
VI.2.2. Modelado por homología de α - y β -tubulina	22
VI.2.2.1. Validación del modelado	23
VI.2.3. Simulación molecular del heterodímero $\alpha\beta$	24
VI.2.3.1. Parametrización	24
VI.2.3.2. Simulación MD del dímero $\alpha\beta$	24
VI.2.4. Simulación del heterotetrámero $\alpha_2\beta_2$	26
VI.2.5. Agrupamiento (<i>clustering</i>) y acoplamiento molecular	28
VI.2.6. Simulación molecular de los complejos ligando-dímero β_2	34
VI.2.6.1. Cambios en la estructura a través del tiempo	34
VI.2.6.2. Puentes de hidrógeno	38
VI.2.6.3. Estructuras secundarias	39
VI.2.6.4. Cálculos de energía de unión	42
(a) Método MM/PBSA	42
(b) Método LIE	42
(c) Cálculo de energía mediante MM/PBSA y LIE	42
VII. DISCUSIÓN DE RESULTADOS	44
VII. 1. Primera parte	44
VII.1.1. Permeabilidad, hidrofobicidad y BHE	44
VII.1.2. Relevancia de los descriptores logP y tPSA	45
VII.1.3. Predicción del transporte a través de la BHE	45
VII. 2. Segunda parte	46
VII.2.1. Modelado de proteínas por homología	46
VII.2.2. Caracterización dinámica del dímero $\alpha\beta$	46
VII.2.3. Caracterización dinámica del heterotetrámero $\alpha_2\beta_2$	47

VII.2.3.1. Heterotetrámero GDP ₂ ,GTP ₂ -α ₂ β ₂	48
VII.2.3.2. Heterotetrámero EPA ₂ -GDP ₂ ,GTP ₂ -α ₂ β ₂	48
VII.2.4. Agrupamiento (<i>clustering</i>) y acoplamiento molecular de compuestos similares a EPA	49
VII.2.5. Suficiencia de β ₂ para la simulación. Necesidad de controles en la simulación	49
VII.2.5.1. Condiciones del equilibrio dinámico	50
VII.2.5.2. Puentes de hidrógeno. Comparación de sistemas con EPA y con control negativo	50
VII.2.5.3. Estructuras secundarias a través del tiempo	51
VII.2.6. Energías de unión. Comparación entre aproximaciones MM/PBSA y LIE	51
VII.2.7. Elección del <i>hit</i> computacional	52
VII.2.7.1. Relaciones estructurales	52
VII.2.7.2. Reacciones de biotransformación	53
VII.2.7.3. Impacto de las interacciones de los ligandos con las isoformas de β-tubulina	55
VIII. CONCLUSIONES	57
IX. PERSPECTIVAS	57
X. ANEXO I	58
X.1. Cribado virtual	58
X.1.1. Coeficiente de Tanimoto	58
X.1.2. Similitud estructural	58
X.1.3. Similitud de interacción proteína-ligando (PLIFs)	59
X.2. Relaciones cuantitativas de estructura-actividad (QSAR) y estructura-propiedad (QSPR)	59
X.2.1. Descriptores moleculares	59
X.2.2. Ecuación de Clark	59
X.2.3. Regresión lineal múltiple	60
X.2.3.1. Colinealidad	61
X.2.3.2. Prueba de significancia de la regresión	61
X.2.3.3. Grados de libertad	61
X.2.3.4. Prueba global de significancia	62
X.2.3.5. Prueba de significancia a coeficientes individuales	65

X.2.4. Adecuación del modelo	66
X.2.4.1. Residuales estandarizados (d_i)	66
X.2.4.2. Residuales estudentizados (r_i)	66
X.2.4.3. Distribución de los residuales	66
(a) Prueba de Shapiro-Wilk	66
X.2.4.4. Intervalos de confianza para los coeficientes de regresión	67
X.2.4.5. Distancia de Cook	67
X.2.5. Validación de la regresión	68
X.2.5.1. Validación interna	68
XI. ANEXO II	69
XI.1. Modelado de proteínas	69
XI.1.2. Análisis cualitativo del modelo de energía (QMEAN)	70
XI.2. Parametrización	71
XI.3. Acoplamiento molecular	72
XI.3.1. SF basadas en campos de fuerza	72
XI.3.2. SF empíricas o de enfoque bayesiano	72
XI.3.3. SF basadas en el conocimiento o de enfoque frecuentista	72
XI.3.4. SF en programas de acoplamiento molecular	73
XI.3.4.1. SF en AutoDock Vina	73
XI.3.5. Algoritmos genéticos	73
XI.4. Dinámica molecular	75
XI.4.1. Campo de fuerza (FF)	75
XI.4.1.1. CHARMM	75
XI.4.1.2. GROMOS	75
XI.4.2. Minimización de la energía	75
XI.4.2.1. <i>Steepest Descent</i> (SD)	76
XI.4.2.2. Gradiente conjugado (CG)	76
XI.4.3. Ensemble o colectivo canónico (NVT)	76
XI.4.3.1. Método de integración de <i>Leap-frog</i>	77
XI.4.3.2. Distribución de Maxwell-Boltzmann	77
XI.4.3.3. Restricciones holonómicas (L)	78
XI.4.3.4. Cálculo de la temperatura	78
XI.4.4. Ensemble o colectivo isobárico isotérmico (NPT)	78

XI.4.4.1. Presión del sistema	78
XI.4.5. Simulación de dinámica molecular	79
XI.4.5.1. Método de integración de Verlet	79
XI.4.5.2. Intervalo de tiempo de integración	81
XI.4.5.3. Condiciones de la dinámica	81
XI.4.5.4. Temperatura y termostato de Berendsen	81
XI.4.5.5. Presión y barostato de Parrinello-Rahman	81
XI.4.5.6. Límites del sistema	82
XI.4.5.7. Sumas de Ewald y PME	82
XI.4.5.8. Interacciones de van der Waals	83
XI.4.5.9. Interacciones hidrofóbicas	84
XI.4.6. Análisis de MD	84
XI.4.6.1. Agrupamiento o <i>Clustering</i>	84
XI.4.6.2. Raíz de la desviación cuadrática media (RMSD)	85
(a) Equilibrio del sistema	85
XI.4.6.3. Raíz de la fluctuación cuadrática media (RMSF)	86
XI.4.6.4. Radio de giro	86
XI.4.6.5. Puentes de hidrógeno	87
(a) Prueba de Mann-Whitney	87
XI.4.6.6. Análisis de estructura secundaria	88
XI.4.6.7. Matrices de distancias	88
XI.4.6.8. Energía de unión	88
(a) Área superficial de Poisson-Boltzmann (PBSA)	88
(b) Aproximación lineal a la energía de unión (LIE)	90
XII. ANEXO III	91
XII.1. Generación de todos los modelos lineales posibles	91
XII.2. Residuales estudentizados de los seis modelos lineales más significativos	92
XII.3. Simulación MD clásica	95
XII.4. Ligandos estabilizadores acoplados en el heterodímero $\alpha\beta$	95
XII.5 RMSD en el sistema $\alpha_2\beta_2$	97
XII.6. Sistema homodimérico EPA ₂ - β_2	100
XII.6.1. Variables termodinámicas	100
XII.6.2. Equilibrio dinámico	102

XII.6.3. Matriz de distancias mínimas	103
XII.7. Energías de solvatación y energías de unión de los ligandos	103
XIII. REFERENCIAS	105

ÍNDICE DE FIGURAS Y TABLAS

Figura 2.1. Representación de un corte coronal encefálico	4
Figura 2.2	5
Figura 2.3. Procesos y vías de señalización implicados con A β 42	6
Figura 2.4. Estructura de los microtúbulos	7
Figura 2.5. Sitios de unión de los desestabilizadores y estabilizadores en el dímero de α , β -tubulina	9
Figura 2.6. Estructura molecular de algunos estabilizadores moleculares	11
Figura 5.1. Metodología general empleada	13
Figura 6.1. Representación de estructuras terciarias de modelos	22
Figura 6.2. Valores z_c obtenidos para los modelos construídos	23
Figura 6.3. Validación de los modelos de tubulinas	24
Figura 6.4. Representación tridimensional de la estructura del heterodímero $\alpha\beta$	25
Figura 6.5. Desplazamientos del dímero $\alpha\beta$ a través del tiempo	25
Figura 6.6. Estructura del heterotetrámero $\alpha_2\beta_2$	26
Figura 6.7. Cálculo de propiedades de los sistemas $\alpha_2\beta_2$	27
Figura 6.8. Estructuras de los ligandos acoplados con PLIF > 0.5	29
Figura 6.9. Gráfico de valores de SF (kcal/mol) en cada acoplamiento	30
Figura 6.10. Imágenes de los acoplamientos moleculares	32
Figura 6.11. Subespacios farmacofóricos representados con esferas, definidos por PLIFs	33
Figura 6.12. Gráficos de RMSD de los sistemas	35
Figura 6.13. Gráficos de RMSD de los ligandos	36
Figura 6.14. Gráficos de RMSF de los sistemas	37
Figura 6.15. Gráficos de radios de giro de los sistemas	38
Figura 6.16. Diagramas de cajas del número de puentes de hidrógeno formados cada 2 ps en los diferentes sistemas	39
Figura 6.17. Asignación de estructuras secundarias a través del tiempo para las cadenas B	40
Figura 6.18. Asignación de estructuras secundarias a través del tiempo para las cadenas D	41
Figura 6.19. Comparación de los métodos de cálculo de energía MM/PBSA y LIE con un diagrama de dispersión	42
Figura 6.20. Comparación pareada de los métodos de cálculo de energía MM/PBSA y LIE	43
Figura 7.1. Modelo del homodímero β_2 representado en estructura terciaria y superficie de van de Waals	47

Figura 7.2. Predicción de las reacciones de biotransformación para el ligando con CID: 59097730	53
Figura 7.3. Predicción de las reacciones de biotransformación para el ligando con CID: 59074402	54
Figura 7.4. Predicción de las reacciones de biotransformación para el ligando con CID: 59958095	54
Figura 7.5. Predicción de las reacciones de biotransformación para el ligando con CID: 69701800	55
Figura 7.6. Alineamiento de las estructuras primarias de las isoformas de β -tubulina	56
Figura 7.7. Estructuras de los <i>hits</i> computacionales resultantes del análisis	56
Figura 10.1. Representación geométrica de los grados de libertad	62
Figura 10.2. Densidad de probabilidad F de Fisher con $v_R=6$ y $v_{res}=2$ grados de libertad	64
Figura 10.3. Densidad de probabilidad t de Student con $v=1$ grado de libertad	65
Figura 11.1 Densidad aproximadamente normal generada a partir del grupo de estructuras proteicas resueltas, con la cual se contrasta el valor z_c	69
Figura 11.2. Ángulos diedros	70
Figura 11.3. Generación de mutaciones mediante el algoritmo genético lamarckiano	75
Figura 11.4. Gráfica de la energía potencial de atracción y repulsión, y su aproximación por la función de Lennard-Jones (L-J)	84
Figura 11.5. Ciclo termodinámico en el que se basa la función de energía libre	90
Figura 12.1. Densidades de los residuales estudentizados utilizados para la detección de valores atípicos de los modelos lineales QSPR	92
Figura 12.2. Gráfico de cuantiles normales para los residuales estudentizados de los modelos lineales QSPR	93
Figura 12.3. Gráfico de valor predicho contra valor observado (experimental) de logBB para el modelo con dos descriptores	94
Figura 12.4. Gráfico de caja para los residuales del modelo final con dos descriptores	94
Figura 12.5. Gráficos de RMSD de las cadenas principales de los polipéptidos en función del tiempo para los sistemas heterodiméricos $\alpha\beta$	96
Figura 12.6. Gráficos de RMSD de los ligandos en el sitio de acoplamiento en función del tiempo	97
Figura 12.7. Gráficos de RMSF de los polipéptidos en función del tiempo para los sistemas heterodiméricos $\alpha\beta$ con los ligandos	97
Figura 12.8. Gráficos de radio de giro de los polipéptidos en función del tiempo para los sistemas heterodiméricos $\alpha\beta$ con los ligandos	98
Figura 12.9. Polígono de frecuencia del análisis de agrupamiento	99
Figura 12.10. Matriz de RMSD del sitio de unión de EPA con distancia mínima de 0.9 nm entre los centros de masa de Q279 y L361	99

Figura 12.11. Relación del tamaño de <i>cluster</i> con su abundancia	100
Figura 12.12. Variables de estado a través del tiempo	101
Figura 12.13. Identificación de un equilibrio dinámico en el control positivo	102
Figura 12.14. Matriz de distancias mínimas entre residuos del dímero β_2	103
Figura 12.15. Contribuciones de los residuos de aminoácidos a la energía de unión de EPA en el control positivo, por el método MM/PBSA	104
Tabla 2.1. Características clínicas de la EA	4
Tabla 2.2. Estabilizadores microtubulares y sus fuentes de obtención	9
Tabla 2.3. Isoformas de β -tubulina	10
Tabla 5.1. Sistemas a los que se les realizó simulación de dinámica molecular	16
Tabla 6.1. Modelos lineales con uno a seis descriptores	19
Tabla 6.2. Datos de las pruebas de significancia global para los seis modelos	20
Tabla 6.3. Datos de las pruebas de significancia por cada descriptor para los seis modelos	20
Tabla 6.4. Datos estadísticos de validación	21
Tabla 6.5. Porcentaje de similitud entre las estructuras primarias de las isoformas de α - y β -tubulina de <i>Homo sapiens</i> y de <i>B. Taurus</i>	22
Tabla 6.6. Análisis de agrupamiento	28
Tabla 6.7. Acoplamientos en las cadenas B de 31 grupos y sus valores de PLIF	29
Tabla 6.8. Interacciones con los residuos de aminoácidos del sitio de unión (EPA) y acoplamiento en la β -tubulina	31
Tabla 10.1. Hipótesis estadísticas concernientes a las medias	63
Tabla 10.2. Tabla del análisis de varianza	64
Tabla 11.1 Evaluaciones realizadas por QMEAN	71
Tabla 11.2. Condiciones de la minimización de energía del sistema	76
Tabla 11.3. Condiciones de NVT	77
Tabla 11.4. Condiciones de NPT	79
Tabla 11.5. Condiciones del método MMPBSA	89
Tabla 12.1. Instrucciones básicas para el análisis preliminar de regresión múltiple	91
Tabla 12.2. Características de los residuales estandarizados y estudentizados de los modelos QSPR	93

SÍMBOLOS, MAGNITUDES Y ABREVIATURAS

Magnitudes y unidades

\AA	Ångström ($1 \text{\AA} = 10^{-10} \text{ m}$)
bar	bar ($1 \text{ bar} = 10^5 \text{ Pa}$)
C_i	Concentración total de la especie i
CM_R	Cuadrado medio de la regresión
CM_{res}	Cuadrado medio residual
$C(x)$	Función de densidad de probabilidad de Cauchy
clogP	Logaritmo de la constante de partición calculada
$DFT\{f\}$	Transformada discreta de Fourier de f
$DFT^{-1}\{g\}$	Transformada inversa discreta de Fourier de g
D_i	Distancia de Cook de i
d	Distancia
d_i	Residual estandarizado de i
$EE(\beta_j)$	Error estándar del regresor β_j
E	Energía
$E(x)$	Valor esperado del valor x
E_{angle}	Energía angular
E_{bond}	Energía de enlace
E_{cor}	Término energético de corrección
E_{dihed}	Energía angular diédrica
E_{dir}	Energía del espacio euclídeo
E_{elec}	Energía electrostática
E_{H-bond}	Energía del puente de hidrógeno
E_{imp}	Energía del ángulo diedro impropio
E_k	Energía cinética
E_{L-J}	Energía potencial de Lennard Jones
E_{MM}	Energía de mecánica molecular
E_p	Energía potencial
E_{rec}	Energía del espacio de Fourier
E_{rot}	Energía rotacional
E_{vdW}	Energía de la interacción de van der Waals
e	Carga elemental
e_i	Residual de i

$\text{erf}(x)$	Función error de x
F	Constante de Faraday
\vec{F}_i	Vector fuerza de i
F_c	Estadístico F de Fisher calculado
F_t	Valor F de Fisher tabulado
fs	Femtosegundo
G_{bind}	Energía de Gibbs de unión
G_{cav}	Energía de Gibbs de formación de cavidad en el solvente
G_{pol}	Contribución energética polar de Gibbs
G_{solv}	Energía de Gibbs de solvatación
G_{vdW}	Contribución energética de van der Waals de Gibbs
$G(\vec{r}_i)$	Función de energía de Gibbs
H_{bind}	Entalpía de unión
H_0	Hipótesis nula
H_1	Hipótesis alterna
h_{ii}	Elemento diagonal ii de la matriz de sombrero
IC_{95}	Intervalo de confianza al 95%
i	Número $i = \sqrt{-1}$
K	Kelvin
k	Constante del resorte de Hook
k_B	Constante de Boltzmann
kcal	Kilocaloría (1 kcal = 4.184 kJ)
kJ	Kilojoule
\log_{BB}	Logaritmo del coeficiente de partición encéfalo-sangre
\log_{BB_i}	Observación i -ésima de \log_{BB}
$\log\hat{\text{BB}}$	Predicción i -ésima de \log_{BB}
$\log P$	Logaritmo de la constante de reparto
M_i	Masa molecular de i
MJ	MegaJoule
m_i	Masa de i
mol	Mol
N	Tamaño poblacional
N_{df}	Número total de grados de libertad
n_i	Tamaño muestral de i

n_t	Número de pasos de tiempo
nm	Nanómetro
ns	Nanosegundo
P	Tensor presión
P	Presión
PRESS	Error de la suma de cuadrados de la predicción
p	Valor p de probabilidad
p_i	Probabilidad de i
PSA	Área superficial polar
q_{int}^2	Coefficiente de validación cruzada interna
q_i	Carga eléctrica de i
RMSD	Raíz de la desviación cuadrática media
RMSF	Raíz de la fluctuación cuadrática media
r	Coefficiente de correlación
r^2	Coefficiente de determinación
\bar{r}^2	Coefficiente de determinación ajustado
r_o^2	Coefficiente de determinación al origen de predicción y observación
$r_o'^2$	Coefficiente de determinación al origen de observación y predicción
r_{val}^2	Coefficiente de determinación de la validación
\vec{r}_{bar}	Centro de masa
r_g	Radio de giro
r_i	Residual estudentizado de i
r_{ij}	Distancia entre i y j
\vec{r}_i	Vector posición de i
SC _R	Suma de cuadrados de la regresión lineal múltiple
SC _{res}	Suma de cuadrados de los residuales
SC _T	Suma de cuadrados total
SF	Función de evaluación
s_R	Error estándar de la regresión lineal múltiple
T	Temperatura
Tan	Coefficiente de Tanimoto
t	Tiempo
tPSA	Área superficial polar topológica
t_c	Estadístico t de Student calculado

t_t	Valor t de Student tabulado
U	Energía interna
U_i	Estadístico de Mann-Whitney
$U(r)$	Función de energía de van der Waals
V	Volumen
V_i	Potencial de i
\vec{v}_i	Vector velocidad de i
W	Estadístico de Shapiro-Wilk
\bar{x}	Media aritmética de x
$x!$	Factorial de x
x_0	Parámetro de Cauchy
z_i	Coefficiente de carga formal de i
z_c	Estadístico z calculado
z_{rk}	Estadístico z por rangos
z_t	Valor z tabulado
α	Área de significancia
β_j	Coefficiente j de la ecuación de regresión lineal múltiple
$\Gamma(x)$	Función gamma de x
γ	Parámetro de Cauchy
$f \Delta g$	Operador diferencia simétrica entre f y g
ΔG	Cambio de la energía de Gibbs
ΔH	Cambio de entalpía
$\delta(x)$	Función delta de Dirac
ϵ	Constante dieléctrica
ϵ_R	Error de la regresión lineal
ϵ_0	Constante de permitividad eléctrica en el vacío
ϵ	Profundidad del pozo de potencial
μ	Media poblacional
ν	Grados de libertad de la función de distribución t de Student
ν_R	Grados de libertad de la regresión lineal múltiple
ν_{res}	Grados de libertad de los residuales
Ξ	Tensor virial
Ξ	Virial
π	Número pi

σ	Desviación estándar poblacional
τ	Constante de tiempo de NVT
τ_P	Constante de tiempo de NPT
ϕ	Ángulo diedro formado por los átomos C-N-C α -C
$\phi(x)$	Función de densidad de probabilidad normal
$\Phi(r)$	Función potencial
ψ	Ángulo diedro formado por los átomos N-C α -C-N
ω	Frecuencia
∇f	Gradiente de f
$\nabla^2 f$	Laplaciano de f
∂f	Derivada parcial de f
$\ \vec{f}\ $	Magnitud del vector f
$\langle f \rangle$	Función f promedio
$\vec{f} \cdot \vec{g}$	Producto escalar del vector f y el vector g
$\vec{f} \otimes \vec{g}$	Producto tensorial del tensor f y el tensor g

Aminoácidos

Abreviaturas	Símbolos	Nombre
Ala	A	Alanina
Arg	R	Arginina
Asn	N	Asparagina
Asp	D	Ácido aspártico
Cys	C	Cisteína
Phe	F	Fenilalanina
Gly	G	Glicina
Gln	Q	Glutamina
Glu	E	Ácido glutámico
His	H	Histidina
Ile	I	Isoleucina
Leu	L	Leucina
Lys	K	Lisina
Met	M	Metionina
Pro	P	Prolina
Ser	S	Serina

Tyr	Y	Tirosina
Thr	T	Treonina
Trp	W	Triptófano
Val	V	Valina

Abreviatura o acrónimo	Significado
ABC	Del inglés: <i>ATP-binding cassette</i>
BHE	Barrera hematoencefálica
CHARMM	Campo de fuerza CHARMM (<i>Chemistry at Harvard Macromolecular Mechanics</i>)
EM	Minimización de energía potencial (<i>Energy minimization</i>)
EPA	Epotilona A
FF	Campo de fuerza
GDP	Guanosina-5'-difosfato
GROMOS	Del inglés: <i>Gröningen Molecular Simulation</i>
GTP	Guanosina-5'-trifosfato
LIE	Energía de interacción lineal (<i>Linear interaction energy</i>)
MD	Dinámica molecular (<i>Molecular dynamics</i>)
MLR	Regresión lineal múltiple (<i>Multiple linear regression</i>)
MM	Mecánica molecular (<i>Molecular mechanics</i>)
MMFF	Campo de fuerza MMFF (<i>Merck Molecular Force Field</i>)
MMPBSA	Área superficial de mecánica molecular de Poisson-Boltzmann (<i>Molecular mechanics Poisson-Boltzmann surface area</i>)
NPT	Colectivo o ensamble isobárico-isotérmico

NVT	Colectivo o ensamble isocórico-isotérmico (canónico)
PDB	<i>Protein data bank</i>
QMEAN	Análisis cualitativo del modelo de energía (<i>Qualitative model energy analysis</i>)
QSAR	Relaciones cuantitativas estructura-actividad (<i>Quantitative Structure-activity relationships</i>)
QSPR	Relaciones cuantitativas estructura-propiedad (<i>Quantitative Structure-property relationships</i>)
SMILES	Especificación de introducción lineal molecular simplificada (<i>Simplified Molecular Input Line Entry Specification</i>)
TAX	Paclitaxel

IA. RESUMEN

La inestabilidad microtubular es un fenómeno existente en todos los subtipos conocidos de la enfermedad de Alzheimer (EA), y está asociada a la formación de placas seniles y de ovillos neurofibrilares. Las placas seniles consisten en agregados intersticiales del péptido A β 42, escindido por las enzimas secretasas a partir de la proteína precursora de amiloide (APP), en un estado anormal de plegamiento, así como acúmulos de otras entidades. Los ovillos neurofibrilares son agregados intracelulares de proteínas hiperfosforiladas, principalmente la proteína tau (TP). En las neuronas de un individuo afectado por la EA, el transporte de vesículas sinápticas y proteínas se ve afectado al depender del sistema microtubular, siendo la afectación del sistema colinérgico uno de las más estudiadas. Además del transporte ineficiente de las vesículas sinápticas en neuronas colinérgicas, la señalización de varios receptores se ve afectada a medida que avanza la EA, debido a sus deficiencias locales y las interacciones con APP, A β 42 o TP. La TP es una proteína asociada a microtúbulos que experimenta hiperfosforilación dependiente de la señalización inducida por A β 42 en el exterior, y en este estado la unión de TP hiperfosforilada disminuye la afinidad entre las subunidades de los microtúbulos, promoviendo su despolimerización, y activa vías de señalización específicas.

Existen agentes moduladores de la dinámica de los microtúbulos, y algunos de ellos, como paclitaxel e ixabepilona tienen indicaciones en diversas neoplasias. Después de los taxenos, hay un número creciente de agentes con actividad microtubular. Las epotilonas, aisladas inicialmente de *Sorangium cellulosum*, presentaron estabilización microtubular y sus derivados ya son fármacos aprobados.

Las herramientas computacionales desempeñan un papel importante en el diseño racional de fármacos, y son muy buenas aproximaciones a las propiedades de las moléculas y sus interacciones. La simulación de dinámica molecular, sustentada en la mecánica molecular, reproduce el comportamiento de las biomoléculas en las condiciones en que se encuentran en las células a través del tiempo, y establece las tendencias de los sistemas biomoleculares que pueden ser comparadas con técnicas experimentales y obtener alta similitud en los resultados.

El objetivo de este trabajo de tesis es obtener las propiedades indicadoras de estabilización microtubular de un grupo de estructuras análogas a la epotilona A (EPA), obtenidas por similitud topológica a través de un cribado virtual en bibliotecas de compuestos. Las estructuras obtenidas se analizaron y filtraron según su capacidad predicha para atravesar la barrera hematoencefálica utilizando relaciones cuantitativas de

estructura-propiedad (QSPR). Las estructuras de α - y β -tubulinas se modelaron por homología a partir de estructuras resueltas experimentalmente, se prepararon los sistemas y se aplicaron simulaciones de dinámica molecular. Los resultados se analizaron para determinar la conservación de las estructuras secundaria, terciaria y cuaternaria de las tubulinas en diversos sistemas de microtúbulos. Se obtuvo el sistema mínimo necesario representativo, el heterotetrámero $\alpha_2\beta_2$, y se calcularon las formaciones de puentes de hidrógeno estabilizadores entre subunidades y entre ligando-proteína, interacciones favorables como interacciones hidrofóbicas y puentes salinos, y valores de energía de unión de los ocho ligandos cribados, además de dos controles, el positivo acoplado a EPA y el negativo sin ligando. También se efectuaron predicciones de reacciones de biotransformación de los compuestos *in vivo*. A partir del análisis se identificaron finalmente las tres estructuras con mejor perfil de estabilización de los microtúbulos, denominadas *hits* computacionales.

IB. INTRODUCCIÓN

La inestabilidad microtubular es un fenómeno presente en los subtipos conocidos de la enfermedad de Alzheimer (EA), y que se presenta en otras tauopatías neurodegenerativas. En la EA, el transporte de vesículas sinápticas en las neuronas experimentan un deterioro considerable al depender del sistema citoesquelético microtubular. Es por esto que los agentes estabilizadores microtubulares han demostrado ser una novedosa solución para el tratamiento de este padecimiento, dentro de los cuales se encuentran las epotilonas (moléculas aisladas inicialmente de *Sorangium cellulosum*). Dos de las dificultades más importantes en la producción de estos compuestos con fines de investigación farmacológica son la escasez o inaccesibilidad de los organismos productores y las síntesis orgánicas multipasos, así como la inespecificidad por las diferentes isoformas de tubulina.

En este trabajo de tesis se analizaron las interacciones y la estabilidad de la epotilona A (EPA) y ocho compuestos seleccionados acoplados al sitio de unión de EPA de un modelo del dímero de tubulinas. Los compuestos seleccionados de alta similitud a la EPA, presentaron una alta capacidad de atravesar la barrera hematoencefálica (BHE), predicha mediante el valor del logaritmo de la constante de reparto encéfalo-sangre (\log_{BB}) calculado con un modelo lineal de dos descriptores basado en relaciones cuantitativas estructura-propiedad (QSPR). Para esto, se generaron y analizaron seis modelos lineales

QSPR siguiendo varias metodologías estadísticas con el objetivo de adecuarlos y validarlos.

El modelo del dímero de tubulinas β_2 , se seleccionó del análisis de estabilidad obtenido de simulaciones de dinámica molecular (MD) de tres modelos: el heterodímero longitudinal $\alpha\beta$, el heterotetrámero $\alpha_2\beta_2$ y el homodímero transversal β_2 . Los compuestos seleccionados se acoplaron en el sitio de unión de EPA en el heterotetrámero $\alpha_2\beta_2$ y en el dímero β_2 , el cual fue equivalente para los objetivos a evaluar. Se analizaron los resultados del acoplamiento y se simularon los ocho sistemas proteína-ligando con MD en el complejo homodimérico β_2 , junto con un control positivo con EPA y un control negativo sin ligando. De las simulaciones se analizaron las desviaciones cuadráticas medias (RMSD) de polipéptidos y ligandos, fluctuaciones cuadráticas medias (RMSF), radios de giro (r_g), formación de puentes de hidrógeno entre subunidades, estructuras secundarias a través del tiempo y estimación de energías de unión por los métodos MM/PBSA y LIE.

A los ligandos con mayor potencial de estabilización microtubular se les realizó análisis de huellas digitales proteína-ligando (*PLIFs*), proposición de subespacios farmacofóricos y predicción de reacciones de biotransformación.

II. ANTECEDENTES

II. 1. Generalidades

La enfermedad de Alzheimer (EA) es un padecimiento neurodegenerativo y la primera causa de demencia en las personas mayores; se estima que en 2050 habrá 13.8 millones de casos a nivel mundial.¹ La EA puede ser de inicio tardío o temprano, este último está asociado a polimorfismos o mutaciones genéticas (Tabla 2.1)². El diagnóstico definitivo de la EA se realiza mediante hallazgos en la necropsia, aunque se investigan varios biomarcadores que puedan revelar el inicio de la enfermedad (Figura 2.1).³

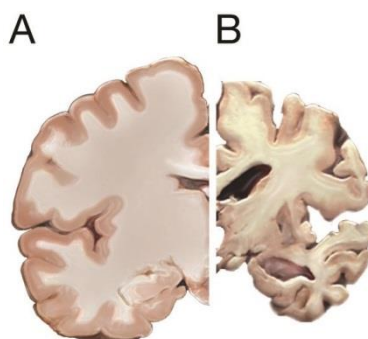


Figura 2.1. Representación de un corte coronal del encéfalo. A. Individuo sano. B. Individuo con EA avanzada (Modificado de *National Institute on Aging/National Institutes of Health*).⁴

Tabla 2.1. Características clínicas de la EA

Característica	EA de inicio temprano	EA de inicio tardío
Edad de inicio	Antes de los 40 años	Después de los 40 años
Amnesia a corto plazo	Avance rápido	Avance progresivo
Amnesia a largo plazo	Aparición rápida	Aparición tardía
Deterioro cognitivo	Establecimiento en enfermedad declarada. El diagnóstico diferencial debe excluir otras enfermedades neurodegenerativas	Establecimiento precoz (hasta 10 años antes de la presentación de los síntomas demenciales)
Tratamiento	Muy limitado en todas las etapas	Limitado a etapas iniciales

II. 2. Etiología

La EA posee características fisiopatológicas distintivas, como las placas seniles, los ovillos neurofibrilares y el deterioro de la neurotransmisión y transducción colinérgica.^{1,2,5}

II. 2. 1. Placas seniles

Las placas seniles son agregados de oligómeros del polipéptido β -amiloide 1-42 ($A\beta_{42}$) que se depositan principalmente en el espacio interneuronal. Los péptidos amiloides son generados a partir de una sucesión de cortes proteolíticos de la proteína precursora de amiloide (APP), siendo los últimos realizados por la β -secretasa y el complejo de la γ -secretasa.⁶ La APP se localiza en balsas lipídicas abundantes en colesterol y fosfolípidos saturados, asociada a varias proteínas como APBA1 y el receptor nicotínico $nAChR\alpha 7$. Después del corte enzimático realizado por γ -secretasa, el fragmento $A\beta_{42}$ es liberado al espacio extracelular, donde se oligomeriza, teniendo efectos moduladores de receptores postsinápticos y canales iónicos. Además, $A\beta_{42}$ activa varias vías de cinasas dependientes de $nAChR\alpha 7$ y altera la dinámica microtubular (Figura 2.2).⁷⁻¹⁰ Además, la presencia de placas seniles activa las células microgliales, las cuales producen citocinas proinflamatorias y especies reactivas de oxígeno (ROS), que inducen la hiperfosforilación de la proteína τ (τ) al interior de las neuronas y astrocitos. Las proteínas τ hiperfosforiladas desestabilizan el citoesqueleto celular, inducen señalización y forman agregados.

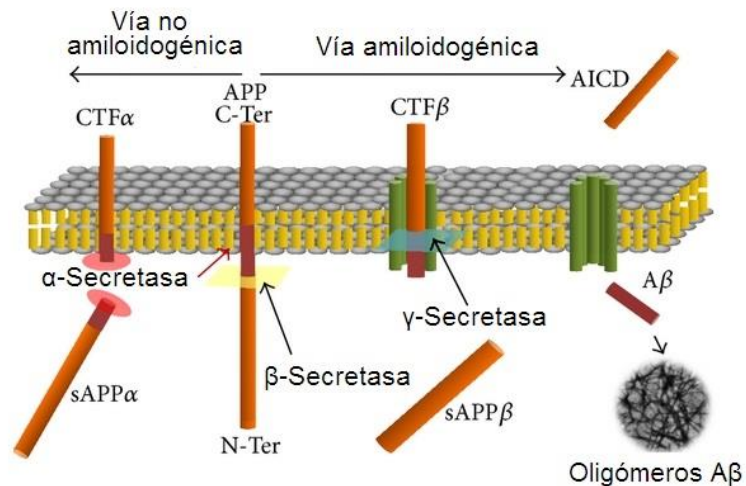


Figura 2.2. Vías de procesamiento de APP. La vía no amiloidogénica genera un fragmento soluble de APP ($sAPP\beta$) y el fragmento C-terminal β ($CTF\beta$) mediante la escisión de la metaloproteína β -secretasa. El procesamiento por la aspartilproteasa β -secretasa produce el fragmento soluble $sAPP\beta$ y el fragmento C-terminal β ($CTF\beta$ o C99). El complejo de γ -secretasa realiza cortes en posiciones variables para liberar al espacio extracelular péptidos amiloides β de diferentes longitudes, y al citoplasma el dominio intracelular de APP (AICD), con actividad señalizadora (Modificado de Picone et al, 2014).¹¹

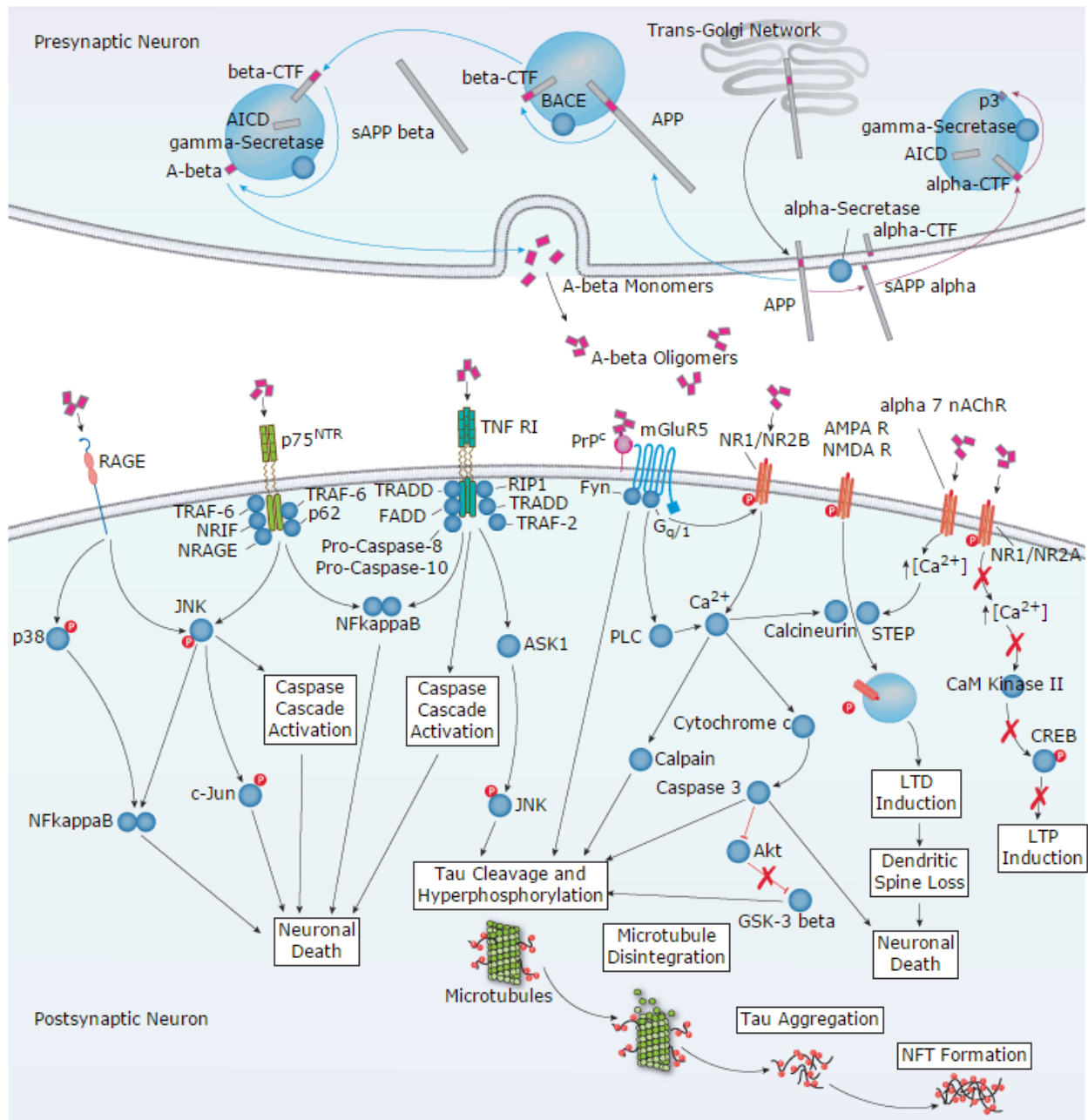


Figura 2.3. Procesos y vías de señalización implicados con Aβ42.¹²

II. 2. 2. Ovillos neurofibrilares

Los ovillos neurofibrilares son agregados de la proteína τ hiperfosforilada que se depositan en el citoplasma de la neuronas.^{1,2,5} La proteína τ es una proteína asociada a microtubulos (MAPs) que en su estado fosforilado ordinario tiene acción estabilizadora microtubular, mientras que en su estado hiperfosforilado es desestabilizadora.⁵ La activación de cinasas dependientes de nAChRα7 por Aβ42 ocasionan dicha hiperfosforilación y, por tanto, la desestabilización microtubular.

II. 3. Microtúbulos

Los microtúbulos (MTs) son estructuras proteicas cuaternarias formadas por copolimerización alternada de α -tubulina y β -tubulina, que constituyen los protofilamentos (Figura 2.3)¹³ o las unidades longitudinales de los MTs. Estas estructuras son una forma común de polímeros de tubulina y constituyen un componente importante del citoesqueleto celular y del huso mitótico. Los MTs consisten de entre nueve y dieciséis protofilamentos. Los MTs participan en múltiples funciones celulares eucariontes, como el tráfico vesicular y organular, la translocación intercompartimental, la asociación con proteínas de andamiaje, las taxias dendríticas, axonales, pseudopódicas, flagelares y ciliares, y en la mitosis y la meiosis.^{10,13-15}



Figura 2.4. Estructura de los MTs. (A) A partir de los heterodímeros $\alpha\beta$ (izquierda) se forman los protofilamentos (centro), los cuales se asocian para formar un MT (derecha). El diámetro de un MT con 13 protofilamentos es de 24 nm, la longitud del anillo dimérico es de 8 nm. (B) Estructura cristalográfica (PDB: 1JFF) del dímero de tubulina. (Modificado de Kollmann, et al.).¹⁵

La estructura de las tubulinas puede dividirse en tres dominios. El dominio amino terminal (del aminoácido 1 al 206) forma el dominio de unión a nucleótido de guanina, y su estructura secundaria contiene láminas β y hélices alternadas. El dominio medio (del aminoácido 207 al 384) forma los contactos entre monómeros a lo largo de los protofilamentos. Finalmente, el tercer dominio está formado por dos hélices largas, el asa que las conecta y el dominio carboxilo terminal.¹³

Los MTs son estructuras dinámicas que experimentan alargamiento y acortamiento de su longitud. Los MTs poseen dos extremos, el extremo terminal de α -tubulina, el extremo (-); y el extremo terminal de β -tubulina, el extremo (+). La demostración que el extremo (+) puede intercambiar GTP, indica que los MTs terminan en una subunidad β . Los protofilamentos experimentan una curvatura en MTs de

rápido crecimiento o que despolimeriza, así como los asociados a análogos de GTP no hidrolizables, lo que indica que la hidrólisis del nucleótido induce un cambio conformacional que causa el encurvamiento de la macroestructura.¹³ La α -tubulina funge como GAP (*GTPase activating protein*) para la β -tubulina adyacente.¹³ La hidrólisis de GTP no ocurre en el último dímero añadido al MT sino en el penúltimo. De esta forma permanece en el extremo (+) un complejo GTP- β -tubulina que asegura la estabilidad del MT naciente. Debido a que el nucleótido se sitúa en la interfase del dímero, se ha propuesto que el encurvamiento del protofilamento durante la despolimerización puede resultar del cambio conformacional de esta interfase.

La inestabilidad dinámica de los MTs consiste en la incorporación de unidades en el protofilamento y su disociación en la misma terminal; rescate y catástrofe respectivamente. El recambio rotatorio ocurre por ganancia de dímeros en un extremo (+) y la pérdida en el otro extremo (-).¹³ Las tubulinas se polimerizan principalmente en los centros de organización de microtúbulos (*Microtubule organizing centers*, MTOCs), donde participa activamente la γ -tubulina.¹⁵

Varios productos naturales ejercen acciones biológicas importantes sobre la estabilidad de los MTs,¹⁶⁻¹⁹ por lo que pueden ser usados como mecanismo de defensa contra otras especies y se clasifican en dos grupos según su acción como: (1) desestabilizadores o (2) estabilizadores microtubulares.

II. 3. 1 Desestabilizadores microtubulares

Los desestabilizadores interactúan con la β -tubulina o en la interfase $\alpha\beta$, favoreciendo la disociación de los MTs.^{13,20,21} Los desestabilizadores microtubulares son citotóxicos a bajas concentraciones y provocan arresto del ciclo celular, más marcado al inicio de la metafase. Algunos de estos agentes tienen uso terapéutico en enfermedades neoplásicas (alcaloides de la vinca, eribulina) o inflamatorias (colchicina) al impactar directamente en la mitosis y en las taxias leucocitarias.^{21,22}

II. 3. 2 Estabilizadores microtubulares

Los estabilizadores (Tabla 2.2) también interactúan con la β -tubulina favoreciendo el rescate, y varios de ellos sólo se unen de manera estable a los protofilamentos y no a los monómeros o heterodímeros aislados. Estos compuestos también son citotóxicos, arrestan el ciclo celular principalmente en anafase y algunos tienen usos terapéuticos en padecimientos oncológicos.^{21,22}

Los estabilizadores más usados comúnmente en la clínica son los taxenos, de los cuales el paclitaxel (TAX) tiene abundantes formas de administración y vehículos; y la ixabepilona, análogo amídico de la epotilona B, los cuales son usados en enfermedades neoplásicas.²⁰⁻²⁶

Tabla 2.2. Estabilizadores microtubulares y sus fuentes de obtención

Tipo de compuesto	Nombre	Fuente
Diterpenoide (taxanos)	Paclitaxel	<i>Taxus brevifolia</i> (“tejo del Pacífico”).
		Endosimbionte <i>Fusarium redolens</i>
		Semisintético
	Docetaxel, tesetaxel, larotaxel, cabazitaxel, ortataxel	Semisintéticos
Sesterterpenoide	Ciclostreptina	<i>Streptomyces</i> sp. 9885
Triterpenoide	Taccalonólidos	<i>Tacca</i> sp.
Policétido	Laulimálido (fijianólido B)	<i>Cacospongia mycofijiensis</i>
	Pelorusido A	<i>Mycale hentscheli</i>
	Epitolonas A, B (patupilona), C, D	<i>Sorangium cellulosum</i>
	Ixabepilona	Sintético
	Sagopilona	Sintético
	(-)-Dictiostatina	<i>Spongia</i> sp.
	Discodermólido	<i>Discodermia dissoluta</i>
	(+)-Dactilólido	<i>Dactylopongia</i> sp.
	(-)-Zampanólido	<i>Fasciospongia</i> sp.
	Alcaloides	Razinilam
Ceratamina A		<i>Pseudoceratina</i> sp.
Ansamicina	Maitansina	<i>Maytenus</i> sp.

Los estabilizadores microtubulares se unen a la β -tubulina en los MTs (Figura 2.5)¹³

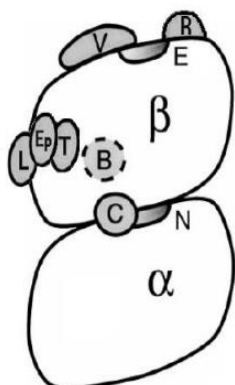


Figura 2.5. Sitios de unión de los desestabilizadores y estabilizadores en el dímero de α, β -tubulina. V: sitio de vincas, R: sitio de rizoxina, L: sitio de laulimálido, Ep: sitio de epotilona, T: sitio de taxenos, B: sitio de bencimidazoles, C: sitio de colchicina (Cano-González, L., 2015).²⁷

Existen numerosas isoformas de tubulinas en un mismo organismo o entre especies. En el humano, se han identificado 10 isoformas β -tubulina, de las cuales 5 se encuentran altamente expresadas en encéfalo (Tabla 2.3).^{14,25}

Tabla 2.3. Isoformas de β -tubulina

Isoforma	Gen	Expresión
β I	TUBB	Ubicua
β IIa	TUBB2A	Predominante en encéfalo
β IIb	TUBB2B	Predominante en encéfalo
β III	TUBB3	Predominante en encéfalo
β IVa	TUBB4A	Predominante en encéfalo
β IVb	TUBB4B	Abundante en corazón y testículos
β V	TUBB6	Ubicua, abundante en corazón
β VI	TUBB1	Abundante en médula ósea y leucocitos
β VII	TUBB7P	Pseudogén
β VIII	TUBB8	Desconocida

El transporte de vesículas sinápticas y las taxias dendríticas y axonales se afectan por la desestabilización microtubular, por lo que la modulación de los MTs es un nuevo enfoque para el tratamiento de la EA mediante el uso de ligandos que estabilicen los MTs.^{1,21,23,25,26,28} Sin embargo, la falta de selectividad a las isoformas de β -tubulina de los moduladores microtubulares usados en la clínica genera múltiples reacciones adversas como mucositis, mielotoxicidad y neurotoxicidad.¹⁴

Las estructuras de varios estabilizadores microtubulares poseen características predominantemente hidrofóbicas (Figura 2.4)²⁹, lo que se relaciona con una baja solubilidad en los compartimentos celulares acuosos y la dificultad de un transporte eficiente a través de las membranas celulares, en particular, hacia el sistema nervioso central (SNC).

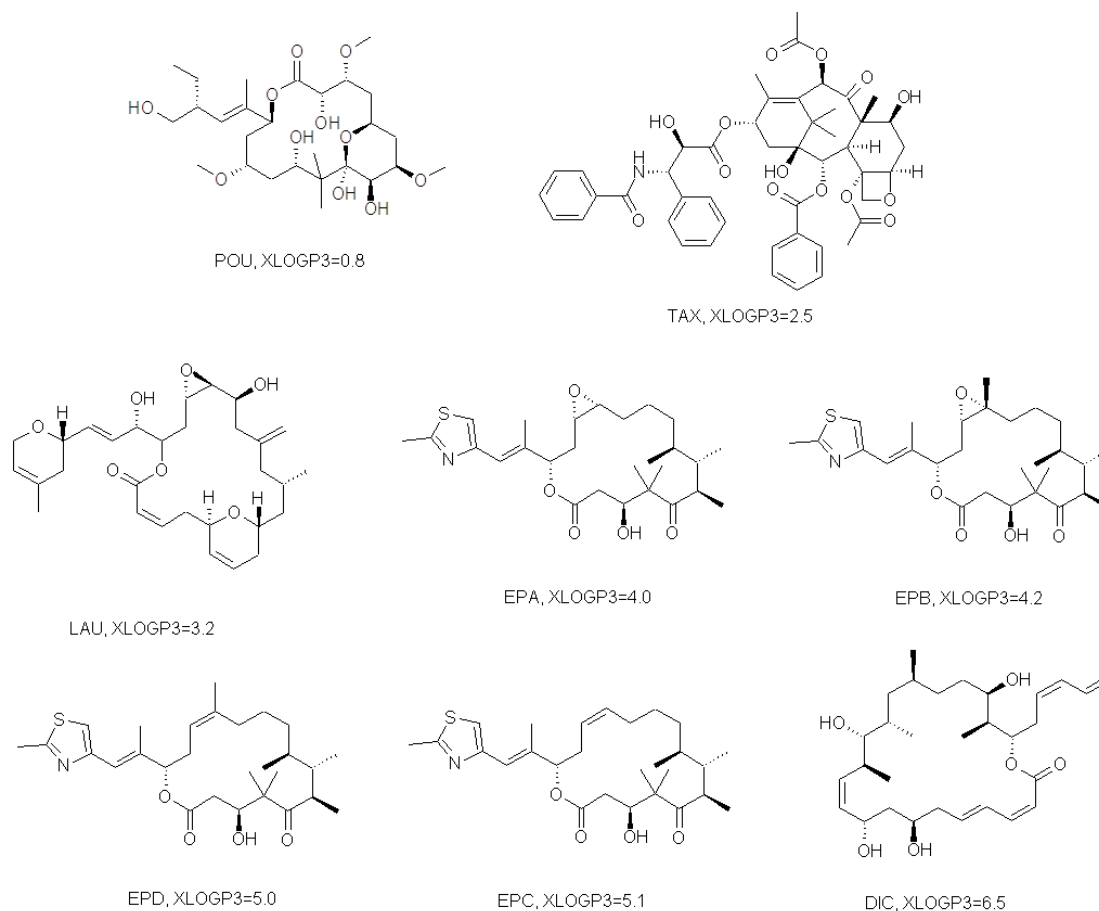


Figura 2.6. Estructura molecular de algunos estabilizadores microtubulares (PubChem Database).²⁹

II. 4. Barrera hematoencefálica

La barrera hematoencefálica (BHE) limita el paso de componentes potencialmente nocivos al SNC. Está conformada por uniones epiteliales estrechas, estabilizadas por cadherinas y expresión abundante de bombas de eflujo. La P-gp/MDR1/ABCB1 (*P-glycoprotein/multidrug resistance protein 1*) es la proteína de eflujo más abundante en las células endoteliales vasculares cerebrales. La P-gp se expresa de manera importante en tejidos epiteliales intestinales y se sobreexpresa en células tumorales. Varios estabilizadores microtubulares son sustratos de P-gp, conduciendo a parte de la resistencia de estas células a los efectos citotóxicos.³⁰

La BHE también tiene menor selectividad hacia los componentes hidrofóbicos, y el tejido encefálico es permeable a los fármacos liposolubles debido a la presencia de gran cantidad de esfingolípidos en las mielinas formadas por los oligodendrocitos.

III. JUSTIFICACIÓN

La EA es una causa importante de discapacidad y demencia en personas de edad avanzada e incluso en algunas personas jóvenes. La EA es compleja y multifactorial, por tanto, su tratamiento debe ser multidisciplinario.

Los MTs son una diana farmacológica relativamente reciente para el tratamiento de la EA, su estabilización por agentes externos ha mostrado ralentizar el deterioro cognitivo asociado con fallas en el transporte vesicular y la axogénesis. Las interacciones entre los estabilizadores microtubulares con la β -tubulina puede estudiarse con herramientas computacionales y la propuesta de un *hit* computacional puede sentar la base para el desarrollo de pruebas biológicas para la evaluación experimental de la actividad de los compuestos.

IV. OBJETIVOS

IV.1. Objetivo general

Evaluar la estabilidad de los complejos de α y β -tubulinas con ligandos estabilizadores microtubulares utilizando herramientas computacionales.

IV.2. Objetivos particulares

- Realizar un cribado virtual en bases de datos para llevar a cabo la búsqueda y evaluación de moléculas similares a EPA con potencial actividad estabilizadora microtubular.
- Proponer un modelo de predicción del transporte a través de la BHE a partir de las características estructurales de los compuestos con similitud a EPA.
- Analizar las interacciones y la estabilidad del ligando microtubular epotilona A (EPA) y ligandos con alta similitud en multímeros de tubulinas empleando herramientas de la simulación molecular, para caracterizar la actividad estabilizadora y las bases de las interacciones proteína-ligando.

V. METODOLOGÍA

El procedimiento general empleado en este trabajo de tesis se esquematiza de en la Figura 5.1. Los Anexos I y II describen de forma detallada los fundamentos de las herramientas computacionales empleadas en este trabajo.

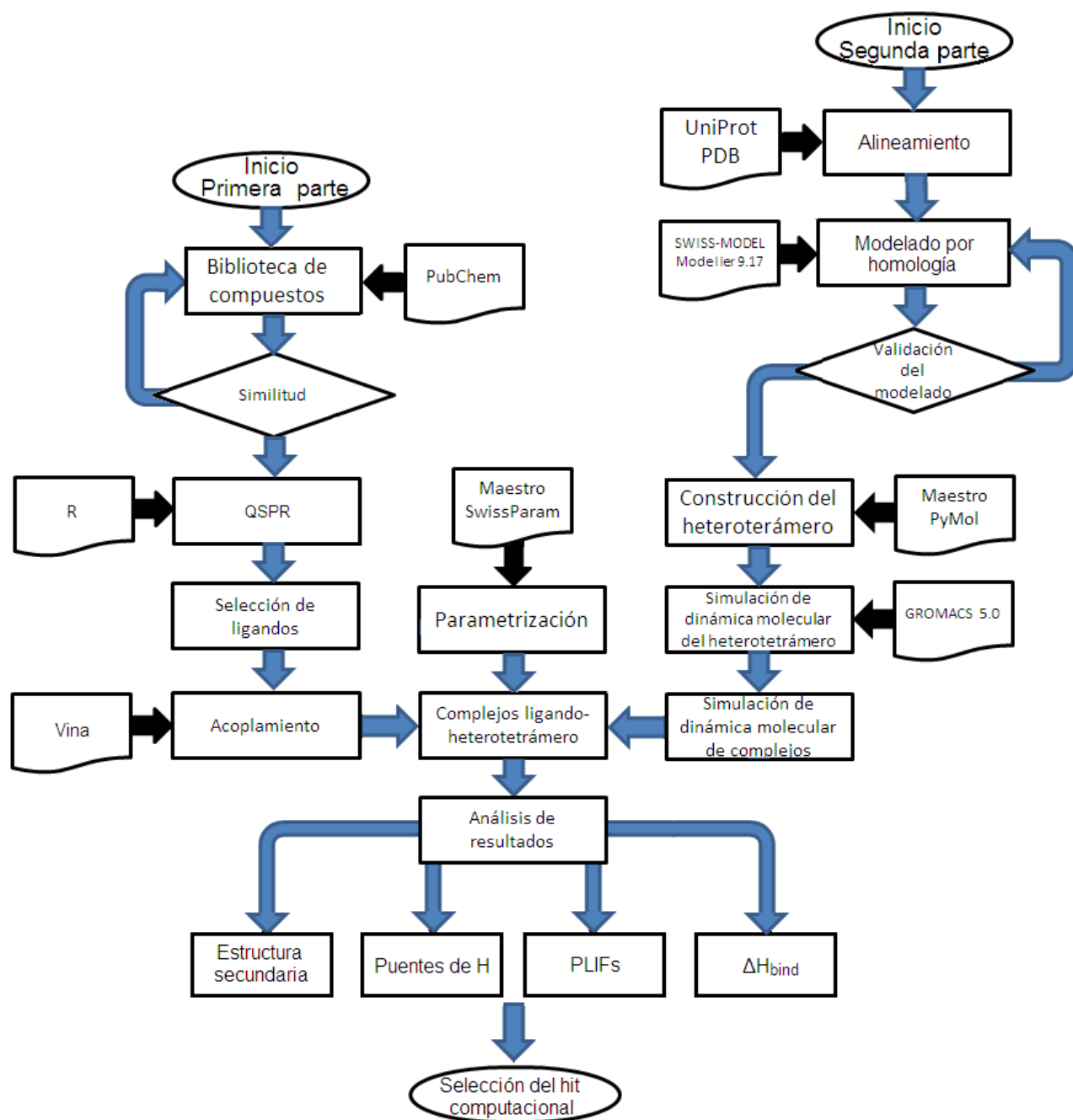


Figura 5.1. Metodología general empleada.

V. 1. Primera parte. Selección de compuestos

V. 1. 1. Similitud molecular

Se filtraron, de la base de datos *PubChem*, los compuestos con 80% de similitud con EPA comparando las huellas digitales binarias de 818 bits con el coeficiente de Tanimoto, un índice que describe la similitud de elementos entre dos conjuntos. Del subconjunto filtrado se generaron SMILES para cada compuesto y, con el servidor *FAF-Drugs3*³¹, se calcularon diecinueve descriptores moleculares fisicoquímicos, estructurales, topológicos y empíricos.

V. 1. 2. Relaciones cuantitativas estructura-propiedad (QSPR)

Para la generación de un modelo lineal basado en la ecuación de Clark,³² la cual es un modelo lineal que calcula el reparto al encéfalo a partir de descriptores moleculares, se utilizaron valores experimentales de la constante de reparto encéfalo-sangre ($\log BB$) de 202 compuestos,³³ a los cuales se les calcularon diecinueve descriptores con el servidor *FAF-Drugs3*³¹. Para este grupo de compuestos se generaron los 524,287 posibles modelos lineales mediante cuadrados mínimos con el lenguaje R³⁴, considerando de uno a diecinueve descriptores con todas sus combinaciones. Para conocer el tamaño del grupo de predicción o de prueba y el grupo de estimación o de entrenamiento, se probaron los 524,287 modelos asignando grupos de prueba de 10%, 15%, 20%, 25%, 30% y 35%, eligiéndose finalmente la proporción 20:80. De todos los modelos generados con la proporción 20:80 se analizaron los estadísticos: coeficiente de determinación (r^2), coeficiente de determinación ajustado (\bar{r}^2), raíz del cuadrado medio residual (CM_{res}), estadístico F de Fisher (F_c) y la probabilidad de buen ajuste debida al azar (p). La validez de los resultados estadísticos de los primeros cinco análisis obtenidos fue confirmada con el programa *BuildQSAR*.³⁵

Para la adecuación de los seis modelos más significativos, se calcularon residuales (e_i), residuales estandarizados (d_i), residuales estudentizados (r_i) y distancias de Cook (D_i) para eliminar valores atípicos. Para los tres primeros métodos, se consideraron los valores atípicos ubicados en las áreas de significancia de 0.05 de su distribución normal, verificada con la prueba de Shapiro-Wilk. Al eliminar los valores atípicos, se realizó la regresión lineal por cuadrados mínimos a los subconjuntos de valores de $\log BB$ y los descriptores para los compuestos, y se analizaron nuevamente los estadísticos r^2 , \bar{r}^2 , F_c , $\sqrt{CM_{res}}$, y p , y los estadísticos de significancia de cada descriptor (t_c).

Para la validación de estos seis modelos se calcularon los estadísticos de validación cruzada interna (q^2_{int}) y los estadísticos de determinación de Tropsha y Golbraikh.³⁶

V.1.3. Predicción del transporte a través de la BHE

A los compuestos con 80% de similitud obtenidos de *PubChem*, se les aplicó el modelo lineal más significativo estadísticamente para filtrar aquellos con valores predichos de logBB que se sabe les confiere capacidad de atravesar la BHE.

V. 2. Segunda Parte. Acoplamiento y simulaciones de dinámica molecular

V. 2. 1. Alineamiento de secuencias

La estructura primaria de la isoforma Ia de α -tubulina (*UniProt*³⁷ Q71U36) y las isoformas IIa, IIb, III y IVa de β -tubulina (*UniProt*: Q13885, Q9BVA1, Q13509 y P04350, respectivamente) de *Homo sapiens* se obtuvieron en formato FASTA, y las secuencias de las estructuras cristalizadas Ib de α -tubulina y IIb de β -tubulina de *Bos Taurus* se obtuvieron de *Protein Data Bank*³⁸ (PDB: 4O4L). El alineamiento se realizó con *Clustal Omega*³⁹ en *UniProt*.

V. 2. 2. Modelado por homología

Las secuencia seleccionadas de α y β -tubulina se modelaron con el servidor *SWISS-MODEL*⁴⁰ y *PROCHECK*.⁴¹ Se seleccionaron los mejores modelos con base en los parámetros de validación que proporciona el servidor *SWISS-MODEL*. Ambos modelos se compararon con el conjunto de estructuras de referencia de alta resolución, diagramas de Ramachandran para analizar estructura secundaria con base en los ángulos diedros de los residuos de aminoácidos, y con el análisis del modelo cualitativo de energía (QMEAN)⁴² para estimar la calidad local de la estructura.

V. 2. 3. Construcción del sistema multimérico

V.2.3.1. Construcción del dímero $\alpha\beta$: Los modelos validados y los cofactores nucleotídicos se alinearon a las estructuras respectivas de PDB 4O4L utilizando *PyMOL*,⁴³ se optimizó la energía del sistema y se resolvieron las estructuras traslapadas con *Maestro 10.5.014*.⁴⁴

V.2.3.2. Construcción del heterotetrámero $\alpha_2\beta_2$: A partir del complejo cofactores-dímero $\alpha\beta$ se construyó el heterotetrámero cofactores- $\alpha_2\beta_2$ con *PyMOL*.⁴³ Se construyó un segundo sistema $\alpha_2\beta_2$ con el ligando de referencia EPA (ligando-cofactores-heterotetrámero).

V. 2. 4. Simulación de dinámica molecular

Los sistemas que se utilizaron para realizar simulaciones de dinámica molecular se muestran en la tabla 5.1.

Tabla 5.1. Sistemas a los que se les realizó simulación de dinámica molecular

Sistema	Ligando
$\alpha_2\beta_2$	(ninguno)
	EPA
β_2	(ninguno)
	EPA
	CID: 59097730
	CID: 59871681
	CID: 60018640
	CID: 59903457
	CID: 59958094
	CID: 59072409
	CID: 59958095
	CID: 69701800

V.2.4.1. Parametrización de ligandos. EPA y los cofactores Mg^{II}-GDP y Mg^{II}-GTP se construyeron y optimizaron en Maestro 10.5.014⁴⁴ y se parametrizaron con el servidor SwissParam⁴⁵ adecuados para ser utilizados con el campo de fuerza CHARMM27.⁴⁶

V.2.4.2. Preparación del sistema. Los sistemas cofactores-proteína se solvataron con agua como solvente explícito, y se adicionaron iones sodio y cloruro para neutralizar y alcanzar una concentración de 0.15 M en GROMACS 5.0.⁴⁷

V.2.4.3. Minimización de energía. La energía potencial del sistema se minimizó empleando el algoritmo Steepest Descent (SD) en GROMACS 5.0.

V.2.4.4. Equilibrio canónico (NVT). Se simuló un sistema cerrado e isocórico con el algoritmo leap-frog, donde se definió la temperatura de referencia (T_0) en 300 K calculada a través de la suma de contribuciones de energía cinética esperada del sistema en GROMACS 5.0 y con restricciones de movimiento en la cadena principal.

V.2.4.5. Equilibrio isobárico (NPT). Se simuló un sistema cerrado isotérmico con el algoritmo leap-frog y se definió la presión de referencia (P_0) de 1 bar, establecida a través del tensor virial (Ξ) del sistema en GROMACS 5.0 con restricciones.

V.2.4.6. Dinámica molecular (MD). El sistema se simuló con MD, con el algoritmo de Verlet durante 50 ns. Las energías electrostáticas de largo alcance se calcularon con el método simplificado de partículas en la malla para las sumas de Ewald (PME) con el potencial coulombico. Las energías de atracción y repulsión se calcularon con el potencial de Lennard-Jones. La temperatura y la presión del sistema se controlaron con el termostato de Berendsen y el barostato de Parrinello-Rahman.

V.2.4.7. Análisis de MD del heterodímero y los heterotetrámeros. Se analizaron los gráficos de raíz de la desviación cuadrática media (RMSD), raíz de las fluctuaciones cuadráticas medias (RMSF) y radio de giro (r_g).

V.2.4.8. Agrupamiento o *clustering*. Se tomaron las conformaciones medias de 31 grupos o *clusters* de $\alpha_2\beta_2$ generados con el método GROMOS, basados en los valores de RMSD con un rango de 0.09 nm, donde la cavidad de unión de EPA tuviera una distancia mínima de 0.9 nm entre centros de masa de dos residuos de aminoácidos usando librerías del programa MDAnalysis.⁴⁸ Para cada uno de los 31 *clusters* se construyó el homodímero β_2 con las cadenas B y D del sistema heterotetramérico.

V. 2. 5. Acoplamiento molecular

Los compuestos filtrados con capacidad de atravesar la BHE se acoplaron en los homodímeros β_2 construidos empleando el programa AutoDock Vina⁴⁹, del cual se analizó la función de evaluación (*Score Function*, SF) y las interacciones proteína-ligando mediante las huellas binarias con el coeficiente de Tanimoto (PLIFs).

V. 2. 6. Simulación MD de complejos proteína-ligando

Los complejos proteína-ligando obtenidos del estudio de acoplamiento se procesaron con los pasos de (a) parametrización, (b) solvatación y neutralización, (c) minimización de energía, (d) equilibrio canónico, (e) equilibrio isobárico, y (f) dinámica molecular. Se incluyeron los sistemas EPA- β_2 como control positivo y β_2 como control negativo.

Se analizaron los gráficos de RMSD, RMSF, r_g , formación de puentes de hidrógeno interfaciales y su comparación estadística de los sistemas con los controles, estructuras secundarias a través del tiempo para las trayectorias de la simulación cada 2 ps, y la estabilidad de los complejos proteína-ligando mediante el cálculo del cambio de entalpía de unión (ΔH_{bind}) usando el método de aproximación lineal de la ecuación de Poisson-Boltzmann (PB) de MM/PBSA⁵⁰ cada 500 ps en el intervalo de tiempo en que se alcanzó el equilibrio dinámico.

Comparación de los ligandos. Se analizaron las estructuras e interacciones de los ligandos con mejores resultados para proponer subespacios farmacofóricos, complementados con la información de Pharmit⁵¹ y se predijeron reacciones de biotransformación con MetaPrint2D-React.⁵²

VI. RESULTADOS

VI. 1. Primera parte. Selección de compuestos

VI. 1. 1. Cribado virtual

De la búsqueda por similitud con el coeficiente de Tanimoto se identificaron 5,163 compuestos con un 80% de similitud con la estructura de EPA, basado en las huellas digitales binarias de 881 bits (ver anexo I, sección X.1).

VI. 1. 2. Relaciones cuantitativas estructura-actividad

Para la predicción del transporte a través de la BHE se generaron 524,287 modelos lineales preliminares a partir de relacionar los 19 descriptores moleculares con el valor experimental de logBB. Los seis modelos más significativos por sus estadísticos se presentan en la Tabla 6.1.

Tabla 6.1. Modelos lineales con uno a seis descriptores

Modelo	r	r ²	\bar{r}^2	s	F _c
logBB=f(tPSA)	0.701	0.491	0.489	0.420	193.1
logBB=f(tPSA, logP)	0.771	0.594	0.590	0.376	145.8
logBB=f(tPSA, logP, MaxSizeRings)	0.781	0.609	0.604	0.370	103.0
logBB=f(tPSA, logP, MaxSizeRings, TotalCharge)	0.787	0.619	0.611	0.366	80.06
logBB=f(tPSA, logP, Rings, MaxSizeRings, TotalCharge)	0.792	0.627	0.617	0.363	65.9
logBB=f(tPSA, logP, Rings, MaxSizeRings, TotalCharge, ratio H/C)	0.795	0.632	0.620	0.362	55.7

Para la exclusión de valores atípicos, se calcularon los residuales, residuales estandarizados (d_i) y residuales estudentizados (r_i) (ver Figura 12.1 y Tabla 12.2 del Anexo III). Los valores de d_i y r_i se compararon a través de sus varianzas y no tuvieron diferencias significativas. Se consideraron valores atípicos cuando $r_i = (-\infty, -2] \cup [2, \infty)$, lo cual constituye aproximadamente el 95% del área bajo la curva normal. A los seis modelos finales se les aplicaron las pruebas de significancia global (Tabla 6.2), pruebas de significancia por descriptor (Tabla 6.3) y validación interna (Tabla 6.4).

Tabla 6.2. Datos de las pruebas de significancia global para los seis modelos^{Nota 1}

Número de descriptores	r^2	\bar{r}^2	F_c	$\sqrt{CM_{res}}$
1	0.6223	0.6198	247.2	0.3344
2	0.8759	0.7672	235.6	0.2614
3	0.877	0.768	157.2	0.2586
4	0.874	0.746	118.2	0.2675
5	0.893	0.798	109.8	0.2469
6	0.897	0.806	94.61	0.2458

Nota 1. Estadísticos de significancia global. r^2 : coeficiente de determinación, \bar{r}^2 coeficiente de determinación ajustado, F_c : estadístico de Fisher, $\sqrt{CM_{res}}$: Raíz cuadrada del cuadrado medio residual (ver anexo I, sección X.2.3).

Tabla 6.3. Datos de las pruebas de significancia por cada descriptor para los seis modelos^{Nota 2}

Descriptor	β_j	$EE(\beta_j)$	t_c	t_t	p
<i>Un descriptor</i>					
tPSA	-0.01158	7.367×10^{-4}	-15.7208	-0.2532	$< 2 \times 10^{-16}$
<i>Dos descriptores</i>					
tPSA	-0.01004	6.044×10^{-4}	-16.6147	-0.2532	$< 2 \times 10^{-16}$
logP	0.1419	0.01434	9.8969	0.2532	$< 2 \times 10^{-16}$
<i>Tres descriptores</i>					
tPSA	-0.01122	6.570×10^{-4}	-17.078	-0.2532	$< 2 \times 10^{-16}$
logP	0.1221	0.01497	8.1570	0.2532	$< 1.65 \times 10^{-13}$
MaxSizeRing	0.01863	0.004665	3.9950	0.2532	4.35×10^{-5}
<i>Cuatro descriptores</i>					
tPSA	-0.01098	6.753×10^{-4}	-16.265	-0.2532	$< 2 \times 10^{-16}$
logP	0.1276	0.01571	8.120	0.2532	$< 1.79 \times 10^{-13}$
MaxSizeRing	0.01513	0.004664	3.243	0.2532	0.00147
TotalCharge	0.04375	0.05341	0.819	0.2532	0.4140
<i>Cinco descriptores</i>					
tPSA	-0.01042	7.830×10^{-4}	-13.303	-0.2532	$< 2 \times 10^{-16}$
logP	0.1470	0.01543	9.525	0.2532	$< 1.79 \times 10^{-13}$
Rings	-0.03522	0.02740	-1.285	-0.2532	0.201
MaxSizeRing	0.02073	0.004582	4.525	0.2532	1.29×10^{-5}
TotalCharge	-0.01448	0.05047	-0.287	-0.2532	0.775
<i>Seis descriptores</i>					

Continuación de la Tabla 6.3

Descriptor	β_j	EE(β_j)	t_c	t_t	p
tPSA	-0.01019	7.597×10^{-4}	-13.420	-0.2532	$< 1.16 \times 10^{-16}$
logP	0.1242	0.01585	7.836	0.2532	$< 1.79 \times 10^{-13}$
Rings	-0.04289	0.02667	-1.608	-0.2532	0.1100
MaxSizeRing	0.01934	0.004666	4.144	0.2532	5.95×10^{-5}
TotalCharge	0.05471	0.05349	1.023	0.2532	0.3082
ratio H/C	-0.1092	0.04730	-2.309	-0.2532	0.0224

Nota 2. β_j : coeficiente del descriptor j, EE(β_j): error estándar del coeficiente β_j , t_c : estadístico t de Student calculado, t_t : valor t de Student tabulado, p: valor de probabilidad de que el resultado del contraste pueda ser igual o más extremo que el obtenido (Ver Anexo I).

Tabla 6.4. Datos de los estadísticos de validación^{Nota 3}

Número de descriptores	r	q_{int}^2	r_{val}^2	r_0^2	$r'_0{}^2$	$\frac{r_{val}^2 - r_0^2}{r_{val}^2}$	k'
1	0.7889	0.6177	0.6223	0.6097	0.6223	0.020	1.00
2	0.8759	0.7634	0.7672	0.7635	0.7672	0.005	1.00
3	0.8767	0.759	0.7685	0.7620	0.7685	0.008	1.00
4	0.8741	0.751	0.7641	0.7593	0.7641	0.006	1.00
5	0.8933	0.783	0.7979	0.794	0.7979	0.005	1.00
6	0.8975	0.788	0.8056	0.8029	0.8051	0.003	1.00

Nota 3: Estadísticos de validación. r: coeficiente de correlación entre valores observados y predichos, q_{int}^2 : estadístico de validación interna, r_{val}^2 : coeficiente de determinación en la validación, r_0^2 : estadístico de determinación ajustado al origen del valor predicho contra el valor observado, $r'_0{}^2$: estadístico de determinación ajustado al origen del valor observado contra el valor predicho, k': pendiente del modelo r_0^2 . Los valores de p en las pruebas de hipótesis para r son menores a 2.2×10^{-16} en los seis casos.

VI. 2. Segunda Parte. Acoplamiento y simulación de dinámica molecular

VI. 2. 1. Alineamiento de secuencias

Las secuencias de aminoácidos de las α - y β -tubulinas de *Homo sapiens* se alinearon con las estructuras primarias de *Bos taurus*, las cuales presentaron una identidad mayor al 90% con esta última (Tabla 6.5).

Tabla 6.5. Porcentaje de similitud entre las estructuras primarias de las isoformas de α - y β -tubulina de *Homo sapiens* y de *B. taurus*

<i>Homo sapiens</i>		<i>B. taurus</i>	Porcentaje de similitud (%)
Gen codificante	Isoforma		
TUBA1A	Ia	TUBA1B	99.56
TUBB2A	IIa	TUBB2A	99.55
TUBB2B	IIb	TUBB2A	100
TUBB3	III	TUBB2A	90.67
TUBB4	IVa	TUBB2A	95.51

La isoforma IA de α -tubulina (TUBA1A, *UniProt*: Q71U36) y la isoforma IIA de β -tubulina (TUBB2A, *UniProt*: Q13885) se seleccionaron para la construcción de los modelos por homología.

VI. 2. 2. Modelado por homología de α - y β -tubulina

El modelado en *SWISS-MODEL* generó respectivamente una estructura terciaria para las secuencias de TUBA1A y TUBB2A (Figura 6.1) a partir de la estructura resuelta de PDB: 4O4L (ver anexo II, sección XI.1).

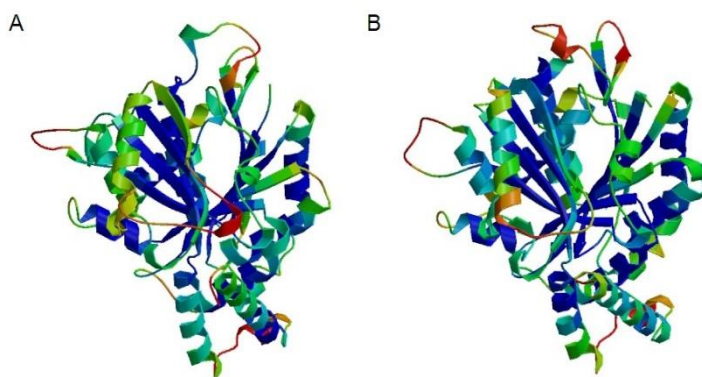


Figura 6.1. Representación de estructuras terciarias de los modelos. (A) TUBA1A. (B) TUBB2A. El código de color muestra las regiones más ordenadas o de menor fluctuación (color azul) a las regiones más desordenadas o de mayor fluctuación (color rojo).

VI.2.2.1. Validación del modelado

Con el servidor *SWISS-MODEL* se calcularon los estadísticos y generaron los diagramas de la validación para cada uno de los modelos. El valor z_c contrasta con la densidad gaussiana de valores z para una biblioteca de estructuras cristalizadas con alta resolución (Figura 6.2).

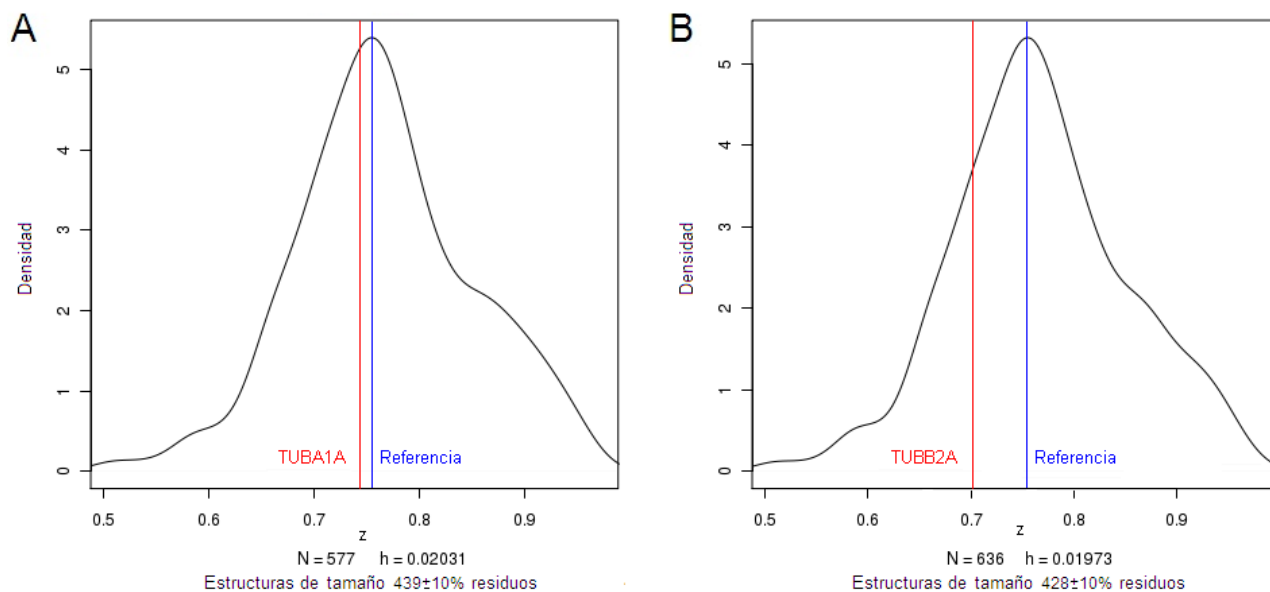


Figura 6.2. Valores z_c obtenidos para los modelos construidos, en línea color rojo. (A) TUBA1A. (B) TUBB2A. Las líneas color azul indican la moda de las densidades obtenidas con regresión Kernel con sus parámetros de ancho de banda h .

Para obtener un análisis del modelo cualitativo de energía (QMEAN), se calcularon los valores z_c para las diferentes comparaciones, (a) interacciones entre $C\beta$, (b) interacciones entre todos los átomos, (c) potencial de solvatación, (d) potencial de torsión, (e) concordancia con la estructura secundaria predicha (SSE) y (f) concordancia con la accesibilidad al solvente (ACC) para TUBA1A y TUBB2A respectivamente (Figura 6.3).

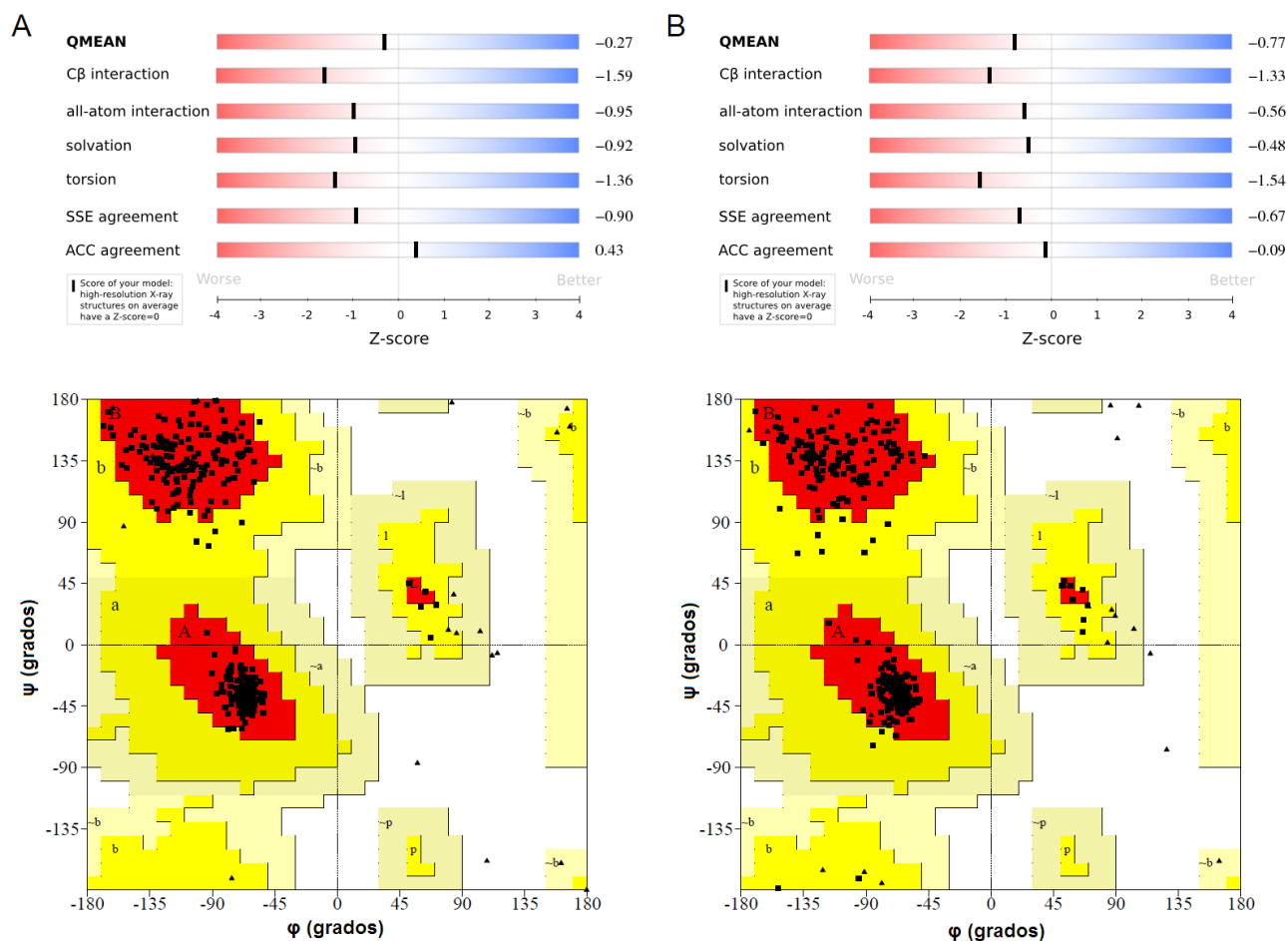


Figura 6.3. Validación de los modelos de tubulinas. QMEAN y diagrama de Ramachandran^{41,53} para (A) TUBA1A y (B) TUBB2A. Los valores z_c se representa como una barra color negro en los rectángulos de colores degradados. Los valores de ψ y ϕ obtenidos para los aminoácidos de los modelos se representan en cuadrados y triángulos (para Gly y Pro) color negro en los diagramas de Ramachandran.

VI. 2. 3. Simulación molecular del heterodímero $\alpha\beta$

VI.2.3.1. Parametrización

Se generaron los archivos de topología para los cofactores (GTP y GDP) y el ligando EPA en el campo de fuerza CHARMM (Ver anexo II, sección XI.2).

VI.2.3.2. Simulación MD del dímero $\alpha\beta$

Para el dímero $\alpha\beta$ con cofactores Mg^{II} -GDP y Mg^{II} -GTP (Figura 6.4) se realizó la simulación de MD durante 50 ns y se calcularon y graficaron el RMSD de la cadena principal y de los cofactores, RMSF y r_g (Figura 6.5, ver Anexo II).

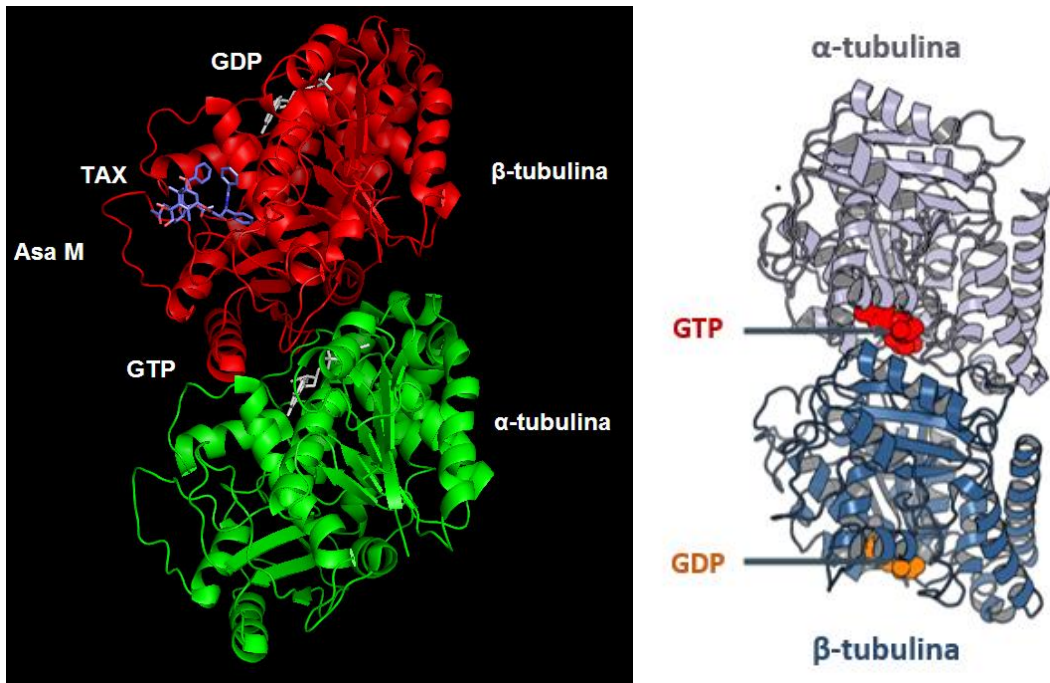


Figura 6.4. Representación tridimensional de la estructura del heterodímero $\alpha\beta$. En color rojo se representa la β -tubulina o cadena B y en color verde la α -tubulina o cadena A. Los nucleótidos se representan en color gris.

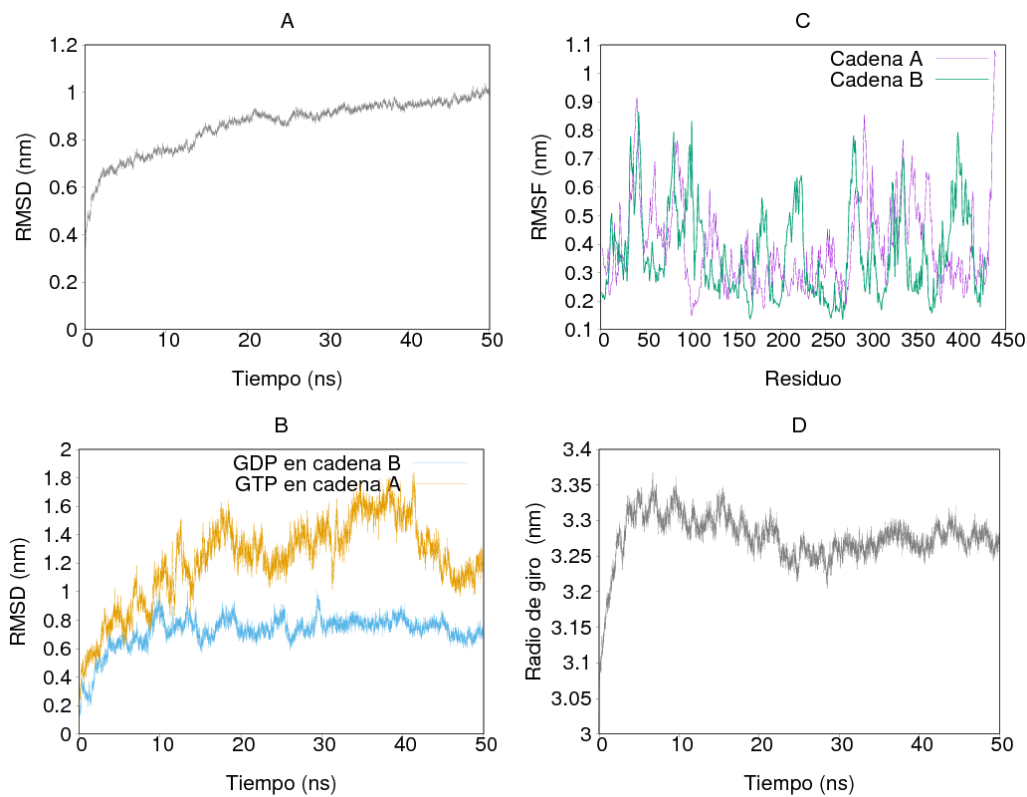


Figura 6.5. Desplazamientos del dímero $\alpha\beta$ en función del tiempo. A. Gráfico de RMSD para la cadena principal de los polipéptidos α y β en el tiempo de simulación. B. Gráfico de RMSD de los cofactores nucleotídicos en sus respectivas cadenas α (A) y β (B). C. Gráfico de RMSF por residuo a través del tiempo. D. Gráfico de radio de giro a través del tiempo.

VI. 2. 4. Simulación del heterotetrámero $\alpha_2\beta_2$

El sistema $\alpha\beta$ no alcanzó un equilibrio dinámico en el tiempo de simulación. Para darle mayor estabilidad al sistema, se modeló el sistema de tubulinas como un heterotetrámero $\alpha_2\beta_2$ (Figura 6.6).

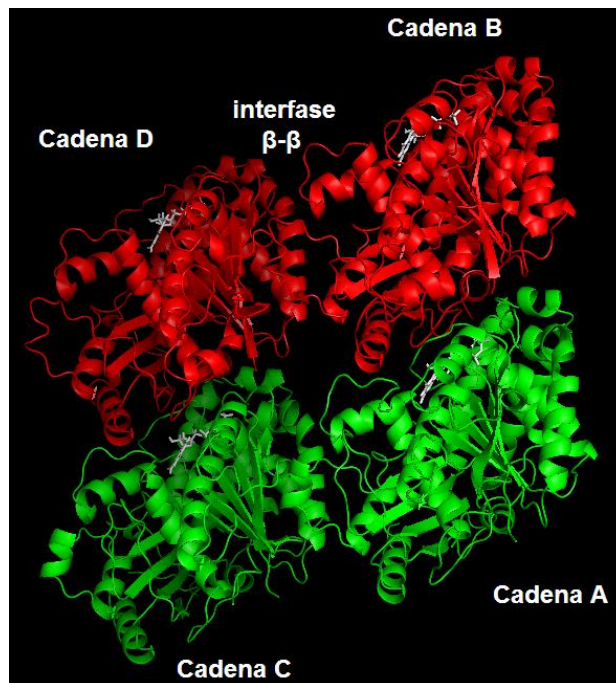


Figura 6.6. Estructura del heterotetrámero $\alpha_2\beta_2$. El asa M de la cadena B se ubica en la interfase β - β y los cofactores GTP en las interfases $\alpha\beta$ respectivas.

De la simulación de los sistemas heterotetraméricos $GDP_2, GTP_2-\alpha_2\beta_2$ y $EPA_2-GDP_2, GTP_2-\alpha_2\beta_2$ se calcularon y graficaron los valores obtenidos de RMSD de las cadenas principales, de los nucleótidos y de EPA, RMSF y r_g como función del tiempo (Figura 6.7).

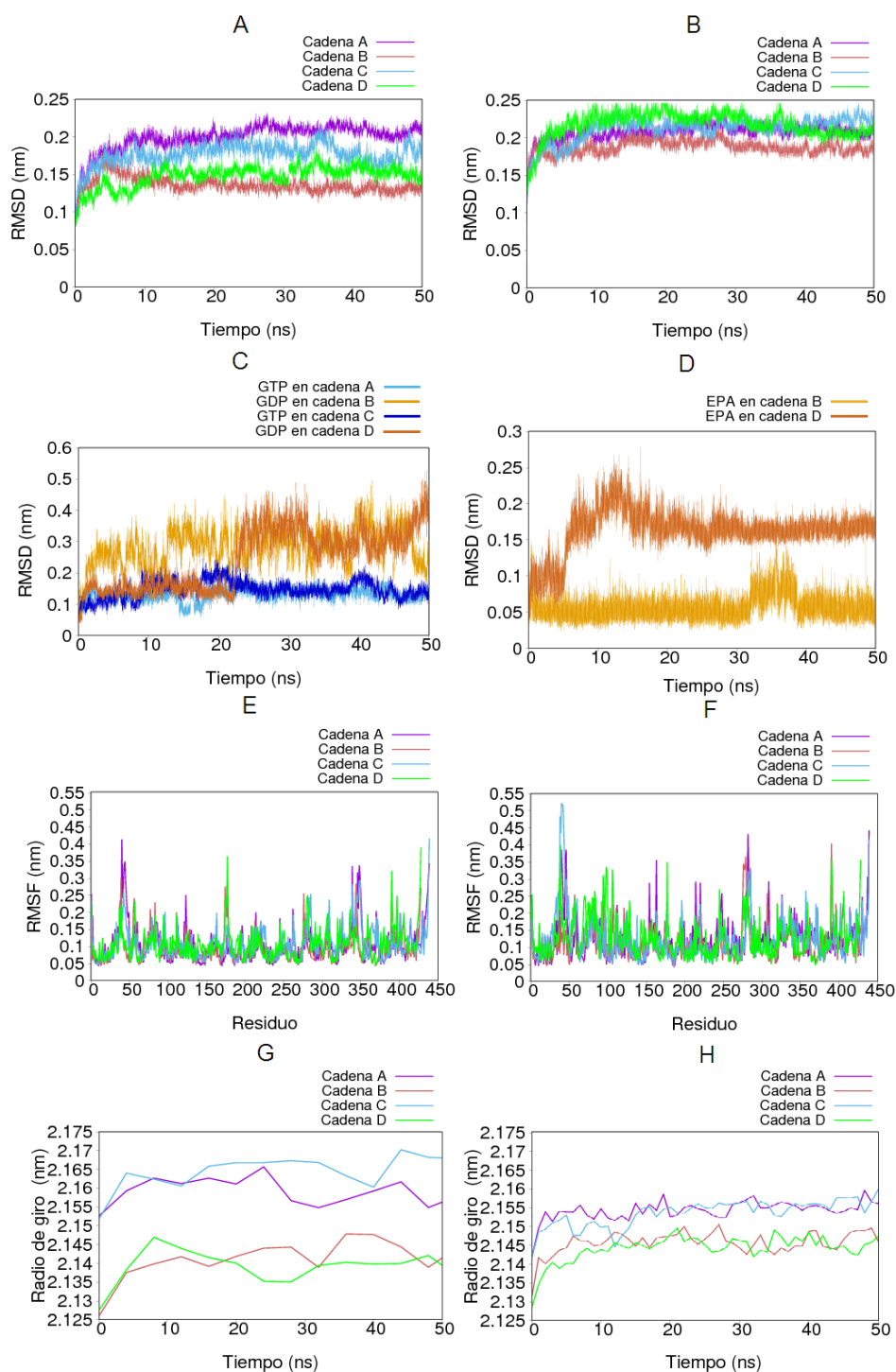


Figura 6.7. Cálculo de propiedades de los sistemas $\alpha_2\beta_2$. Del lado izquierdo se presentan resultados obtenidos para el sistema sin EPA y del lado derecho se muestran los resultados obtenidos para el sistema con EPA. (A) y (B) Muestran las gráficas obtenidas de los valores calculados de RMSD para las cadenas principales de los polipéptidos α (A y C) y β (B y D) a lo largo de la simulación. (C) Gráfico de valores obtenidos de RMSD de los cofactores nucleotídicos en las cadenas A, B, C y D. (D) Gráfico de valores obtenidos de RMSD de EPA en las cadenas B y D. (E) y (F) Gráficos de RMSF por residuo en cada cadena a través del tiempo. (G) y (H) Gráficos de radio de giro a través del tiempo para las cadenas A, B, C y D.

VI. 2. 5. Agrupamiento (*clustering*) y acoplamiento molecular

Se realizó el agrupamiento de aquellas estructuras de la simulación donde el sitio de unión de EPA tenía una distancia mínima de 0.9 nm, entre los centros de masa de Q279 y L361 utilizando librerías de *MDAnalysis*.^{48,54} Para esto se generó una matriz de RMSD (Figura 12.5 del Anexo III) en función del tiempo tomando como límite de diferenciación entre estructuras un valor de 0.09 nm, lo cual permitió generar un total de 31 grupos (Tabla 6.5).

De cada uno de los grupos anteriores, se extrajo la estructura más representativa de las cadenas B del heterotetrámero $\alpha_2\beta_2$. Se realizó el acoplamiento con *AutoDock Vina* (ver anexo II, sección XI.3) de los compuestos filtrados con el modelo de logBB con dos descriptores en las 31 conformaciones medias de B, y se calculó el valor de PLIF (Tabla 6.6). De los 3,007 acoplamientos, se seleccionaron siete conformaciones medias (Tabla 6.7) con base en el valor de PLIF de los ligandos (Figura 6.8) y sus valores de SF (Figura 6.9).

Tabla 6.6. Análisis de agrupamiento. Se muestra el grupo o *cluster* enlistado en orden decreciente del número de conformaciones que lo componen. Se muestran también las conformaciones medias de cada grupo con un valor de PLIF mayor a 0.5 de los ligandos filtrados, con clave CID

Cluster	Tiempo (ps)	Tamaño poblacional	PLIF \geq 0.5	Cluster	Tiempo (ps)	Tamaño poblacional	PLIF \geq 0.5
1	25384	38	0	16	27044	2	0
2	35436	19	0	17	25632	2	0
3	23344	18	0	18	25256	2	0
4	21044	14	0	19	22336	2	0
5	24640	11	0	20	25444	1	0
6	25668	10	0	21	27176	1	1
7	21900	9	0	22	24064	1	0
8	21392	8	0	23	21272	1	1
9	23668	8	0	24	20376	1	0
10	26736	7	0	25	20844	1	1
11	25048	6	0	26	21684	1	0
12	26912	5	1	27	30060	1	0
13	20064	4	0	28	35220	1	1
14	24436	3	0	29	37400	1	3
15	21152	2	0	30	38756	1	0
				31	47756	1	0

Tabla 6.7. Acoplamientos de los ligandos con claves CID en las cadenas B de 31 grupos y sus valores de PLIF

Tiempo de <i>cluster</i> (ps)	CID del compuesto acoplado	PLIF
20844	59097730	0.55
21272	59871681	0.54
26912	60018640	0.50
27176	59903457	0.50
35220	59958094	0.50
	59072409	0.55
37400	59958095	0.54
	69701800	0.50

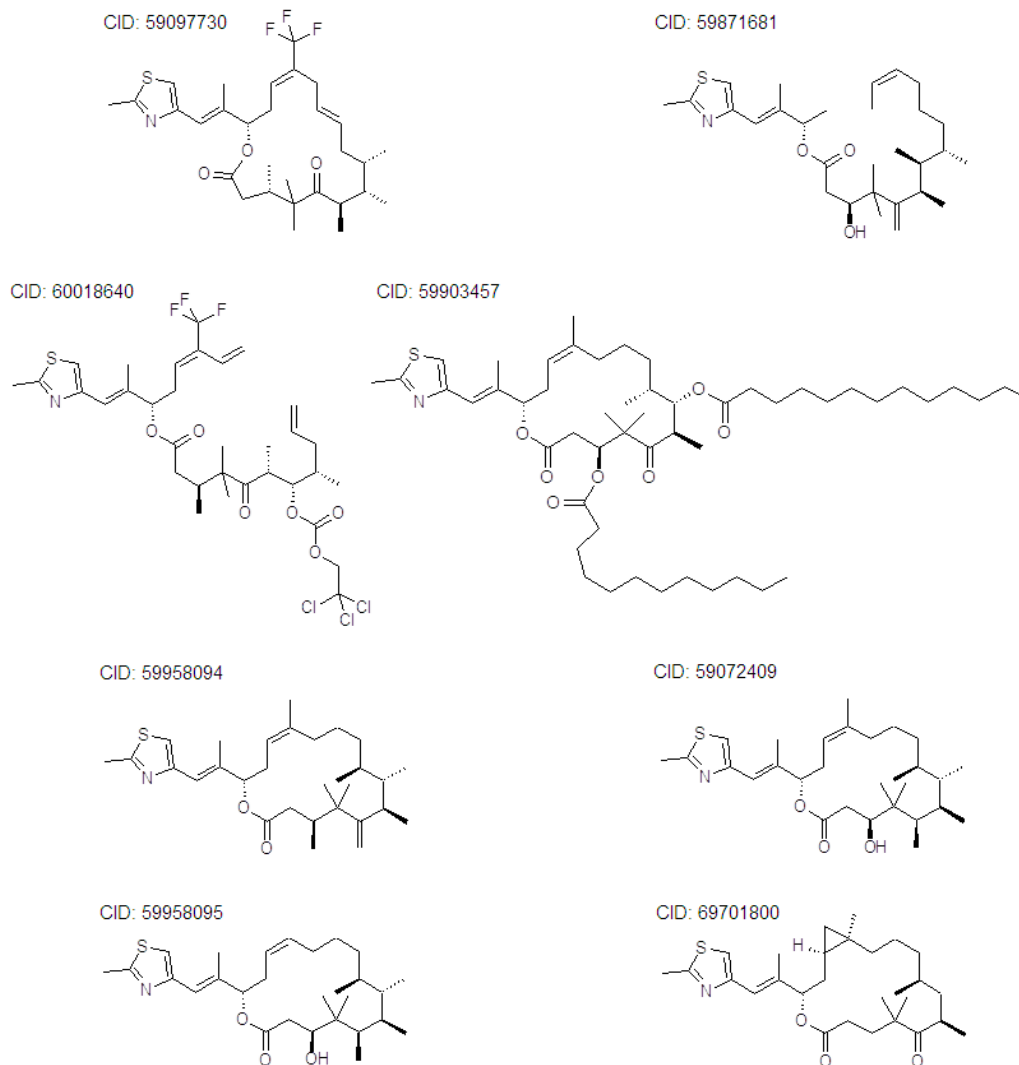


Figura 6.8. Estructuras de los ligandos con clave CID acoplados con PLIF > 0.5.

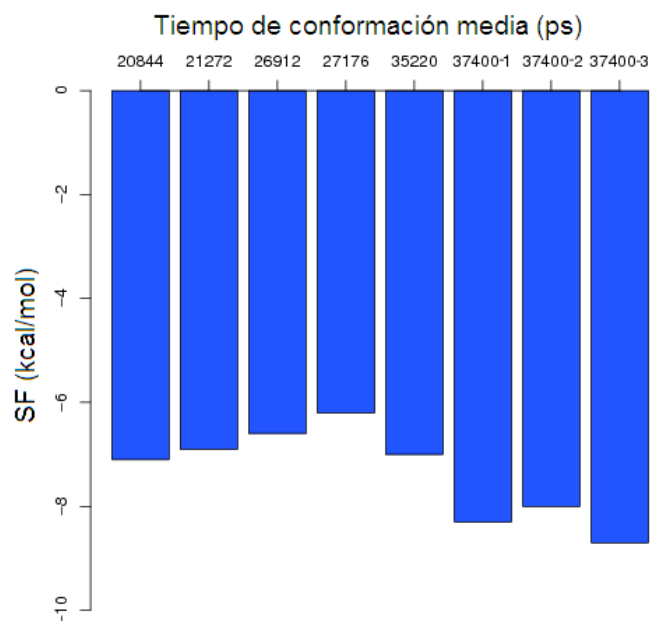


Figura 6.9. Gráfico de valores de SF (kcal/mol) en cada acoplamiento.

Las interacciones de los ligandos en el sitio de acoplamiento de los sistemas β_2 (PLIFs), basados en las distancias respectivas, se muestran en la Tabla 6.7. Las interacciones identificadas de los ligandos con los aminoácidos de la Tabla 6.8 se presentan en la Figura 6.10. Los subespacios farmacofóricos definidos por los PLIFs (Figura 6.11) coinciden con los propuestos por Pharmit.⁵¹

Tabla 6.8. Interacciones con los residuos de aminoácidos del sitio de unión (EPA) y acoplamiento en la β -tubulina. Los residuos de interacción se representan en color verde los compartidos por EPA y el ligando, en color negro los exclusivos de los ligandos, y en color rojo los exclusivos de EPA

Ligando	Tipo de interacción						
	Hidrofóbica				Puente salino	Puente de hidrógeno	Coordinación
59097730 (A)	L215		D224		H227	D224	Q279 (con flúor)
	L228					T274	
				L361		Q279	
59871681 (B)	L215	L217	D224	H227		D224	
	L228		P272			T274	
		Q279		L361		Q279	
60018640 (C)	L215	L217				D224	T274 (con flúor)
	L228	F270	P272			T274	
		Q279	L284	L361		Q279	
59903457 (D)	L215	L217	D224	H227		L273 (cadena principal)	
	L228		P272	T274		D224	T274
				L361		Q279	
59958094 (E)	L215					D224	
	L228		P272	T274		T274	
		Q279		L361		Q279	
59072409 (F)	L215			H227	H227	D224	
	L228		P272	T274		T274	
		Q279		L361		Q279	
59958095 (G)	L215	L217	D224	H227		D224	
	L228		P272	T274		T274	
		Q279	L284	L361		Q279	
69701800 (H)	L215					D224	
	L228		P272			T274	
	R276	Q279	L284	L361		Q279	

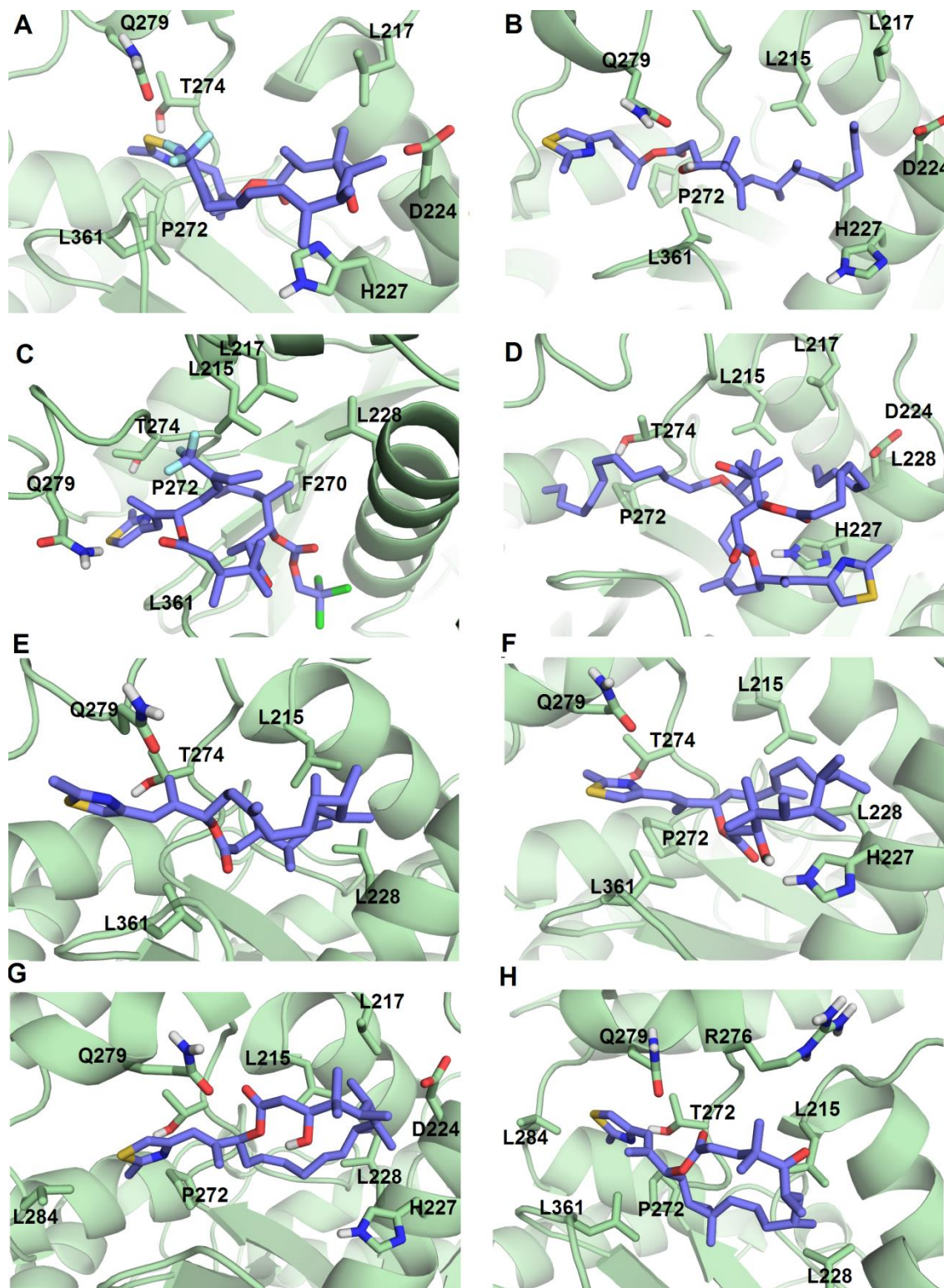


Figura 6.10. Imágenes de los acoplamientos moleculares. Ligandos con CID: (A) 59097730, (B) 59871681, (C) 60018640, (D) 59903457, (E) 59958094, (F) 59072409, (G) 59958095 y (H) 69701800.

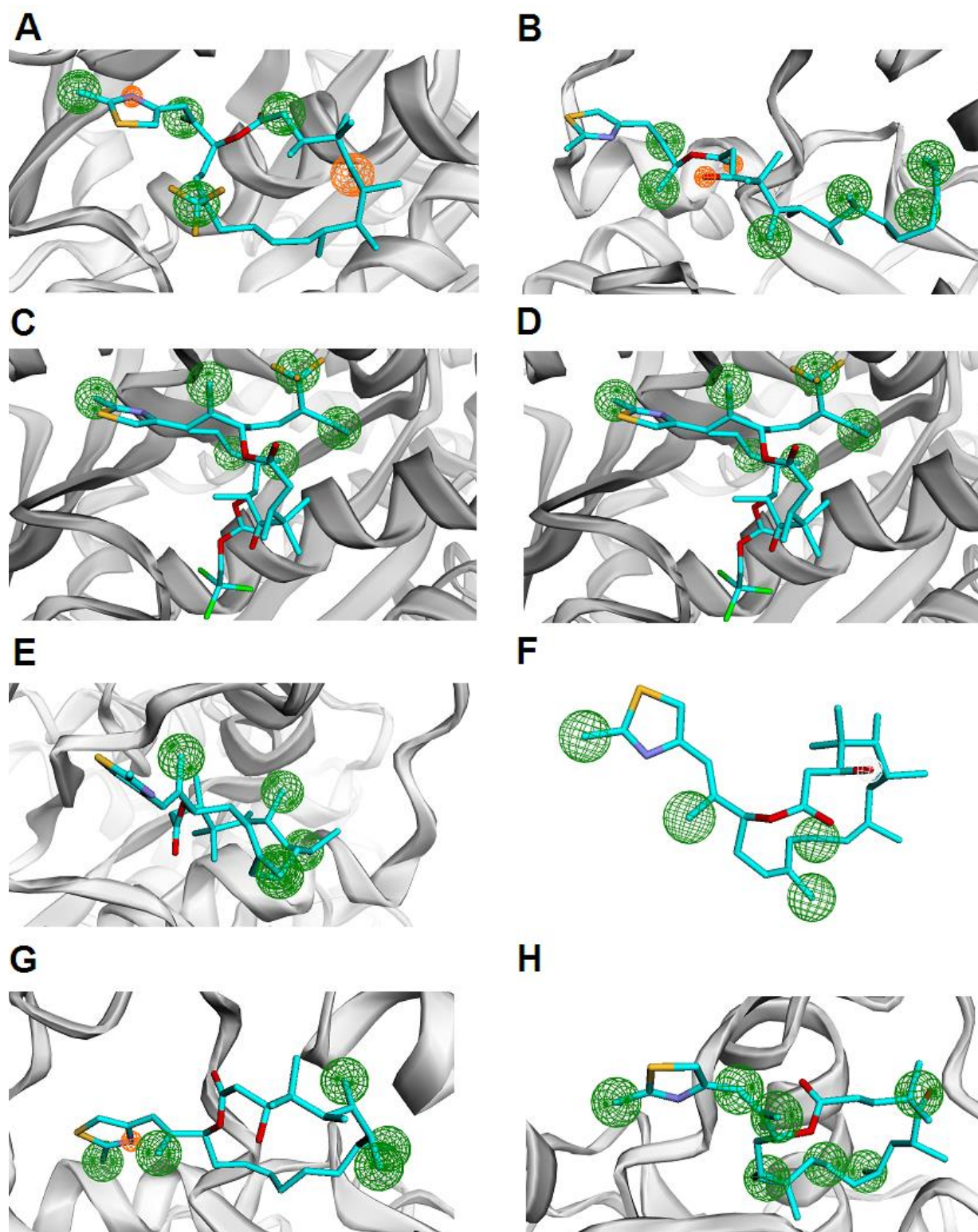


Figura 6.11. Subespacios farmacofóricos representados con esferas y definidos por PLIFs. Ligandos con CID: (A) 59097730, (B) 59871681, (C) 60018640, (D) 59903457, (E) 59958094, (F) 59072409, (G) 59958095 y (H) 69701800.

VI. 2. 6. Simulación molecular de los complejos ligando-dímero β_2

De las conformaciones medias de los seis grupos de la Tabla 6.6, se construyeron dímeros de β_2 con las cadenas B y D, y se realizaron simulaciones de MD de 50 ns a los ocho complejos LIG- β_2 . Adicionalmente, se construyeron el dímero EPA- β_2 como control positivo y el dímero β_2 como control negativo. Para cada sistema se analizó su energía potencial (E_{pot}) durante la minimización, la temperatura (T) en las simulaciones de equilibrio NVT, la presión (P) en las simulaciones de equilibrio NPT, y el volumen (V), la presión (P), y la temperatura (T) durante la producción de MD (ver Anexo II, sección XI.4). Al finalizar la producción se analizaron las desviaciones RMSD, RMSF y r_g ; los puentes de hidrógeno formados, las estructuras secundarias presentes durante la simulación, las matrices de distancia y se calcularon los valores de ΔH de unión de los ligandos.

VI.2.6.1. Cambios en la estructura a través del tiempo

Se graficaron los valores calculados de RMSD como función del tiempo para cada cadena en los homodímeros durante los 50 ns de simulación (Figura 6.12). Las funciones RMSD para las cadenas B (color violeta) y D (color verde) se muestran en una sola gráfica por sistema.

Para los ligandos se graficaron sus funciones RMSD por sistema (Figura 6.13) para las cadenas B (color azul) y D (color amarillo). El ligando en la cadena B es el que se sitúa en la interfase β - β .

Las funciones de RMSF por residuo (Figura 6.14) se calcularon para los intervalos de tiempo de equilibrio dinámico identificados con la aplicación de DFT, para cada sistema.

Se calcularon los gráficos de radio de giro a través del tiempo de simulación (Figura 6.15).

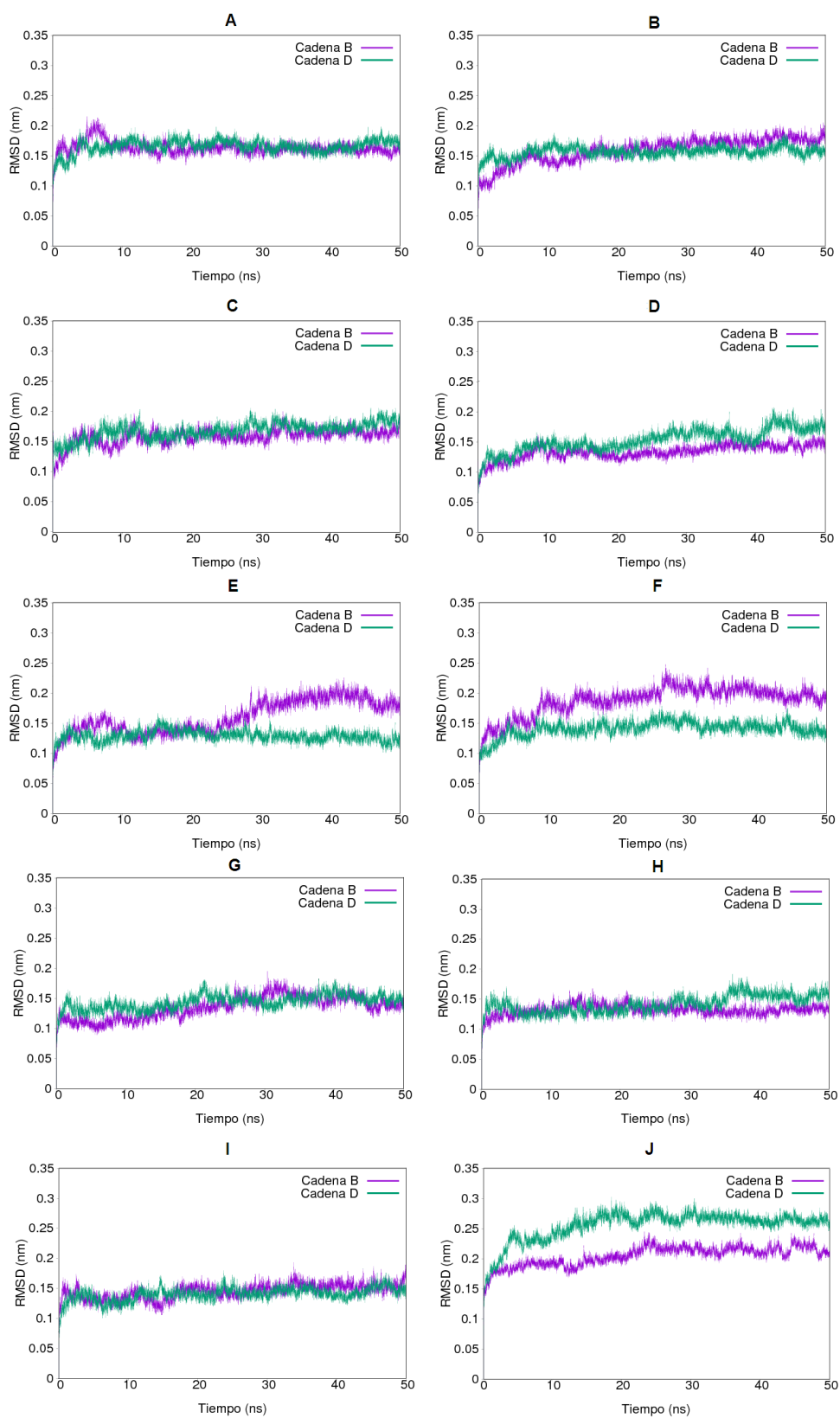


Figura 6.12. Gráficos de RMSD de los sistemas. (A) Control positivo (EPA). Sistemas homodiméricos con los ligandos con CID (B) 59097730. (C) 59871681. (D) 60018640. (E) 59903457. (F) 59958094. (G) 59072409. (H) 59958095, (I) 69701800 y (J) Control negativo.

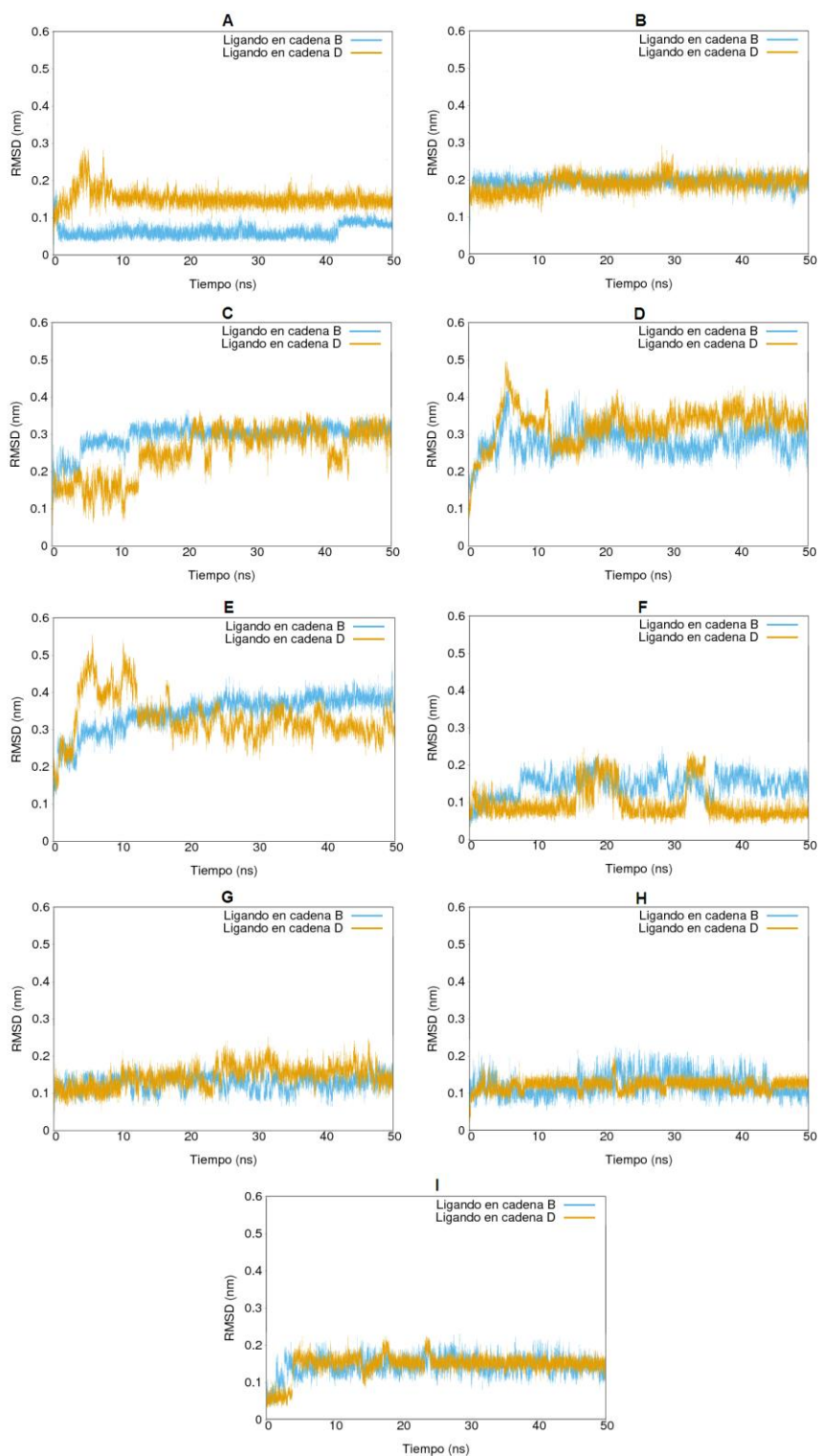


Figura 6.13. Gráficos de RMSD de los ligandos. (A) EPA, (B) 59097730, (C) 59871681, (D) 60018640, (E) 59903457, (F) 59958094, (G) 59072409, (H) 59958095 e (I) 69701800.

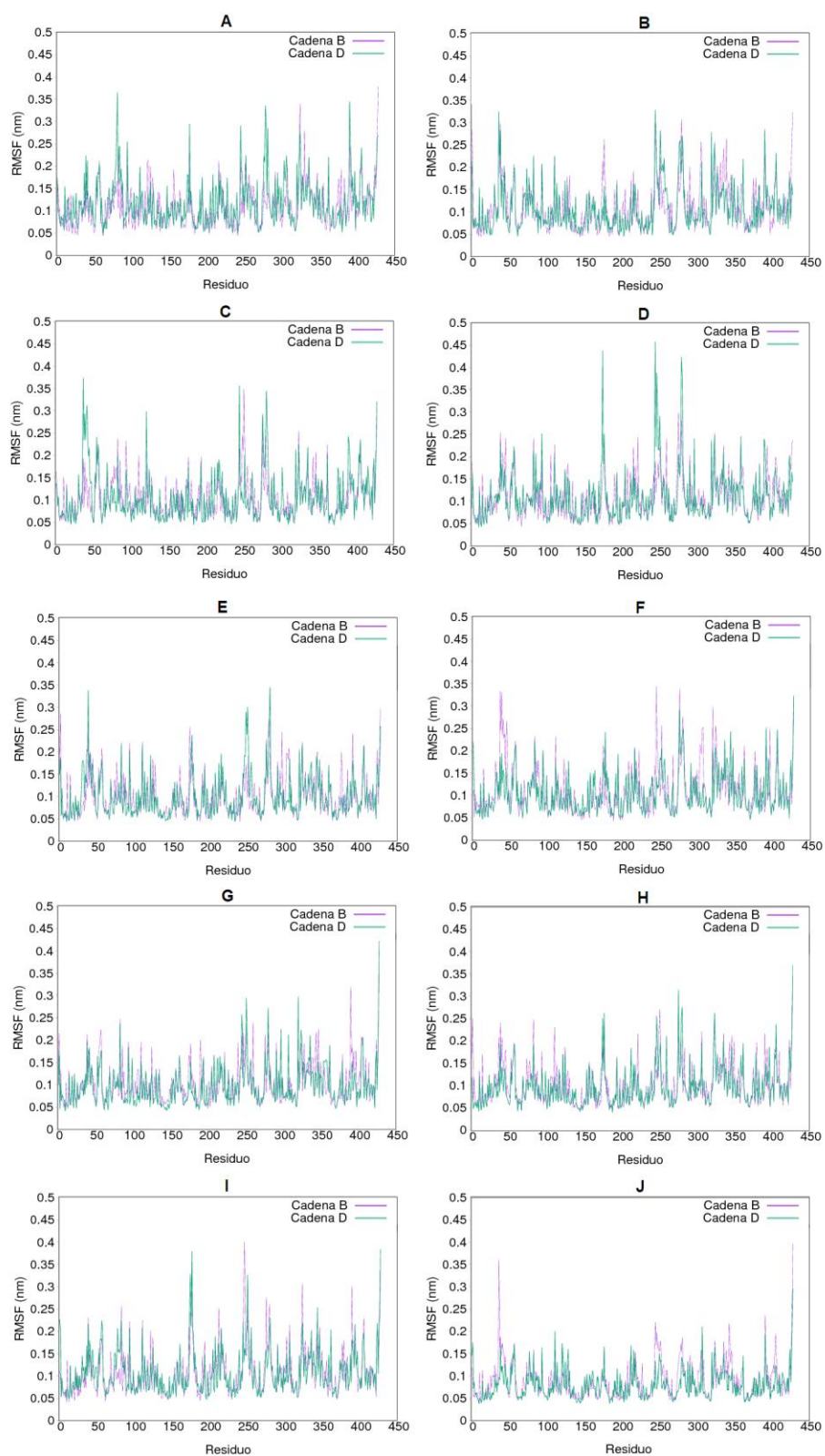


Figura 6.14. Gráficos de RMSF de los sistemas con los ligandos (A) EPA, (B) 59097730, (C) 59871681, (D) 60018640, (E) 59903457, (F) 59958094, (G) 59072409, (H) 59958095 e (I) 69701800. (J) fue el control negativo.

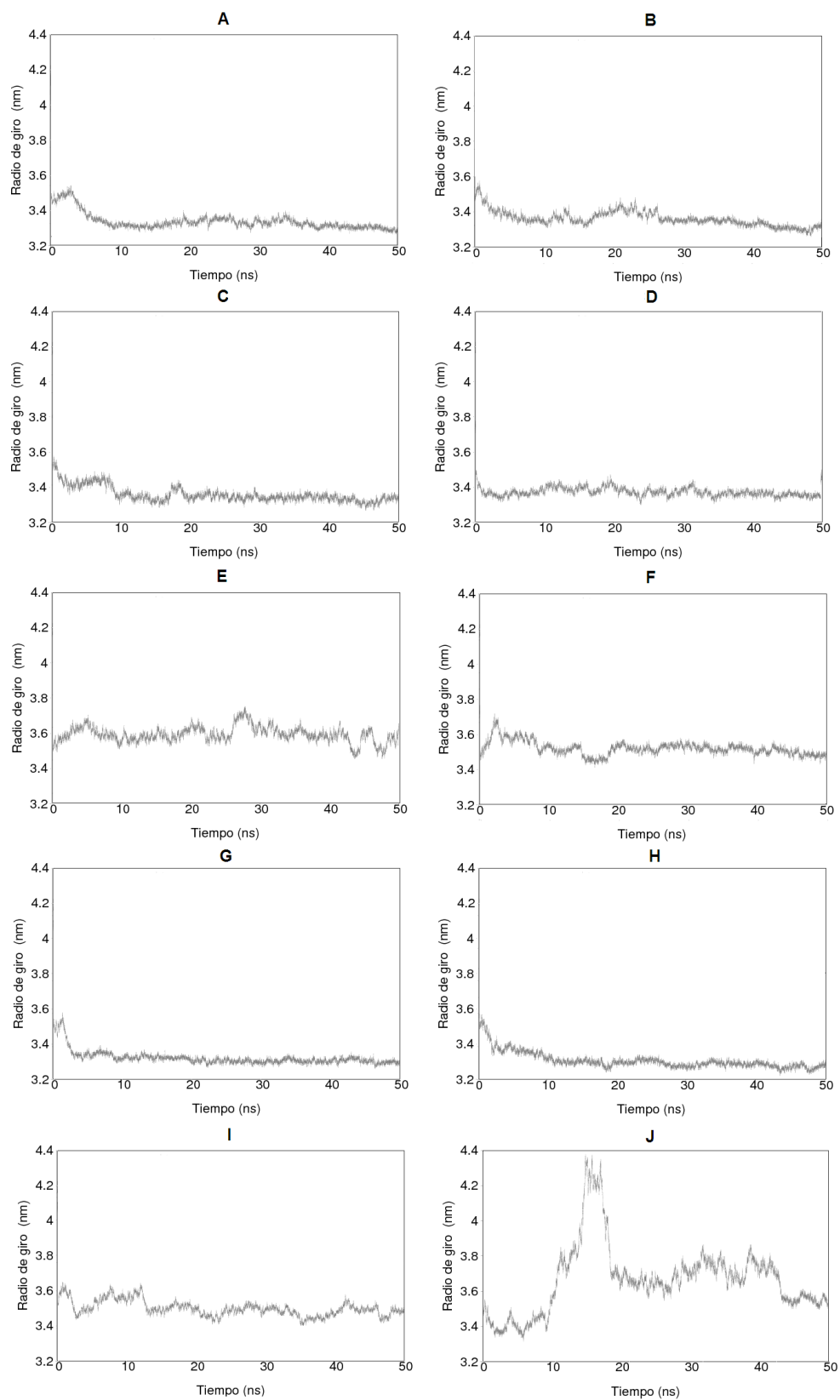


Figura 6.15. Gráficos de radio de giro de los sistemas con los ligandos (A) EPA, (B) 59097730, (C) 59871681, (D) 60018640, (E) 59903457, (F) 59958094, (G) 59072409, (H) 59958095 e (I) 69701800.(J) fue el control negativo.

VI.2.6.2. Puentes de hidrógeno

Se realizó la comparación del número de puentes de hidrógeno en cada sistema (Figura 6.16) mediante la prueba de Mann-Whitney, siendo el compuesto 60018640 estadísticamente equivalente a EPA ($p = 0.2357$). Todos los ligandos, incluyendo EPA, tuvieron una mayor formación de puentes de hidrógeno estadísticamente significativa respecto al control negativo.

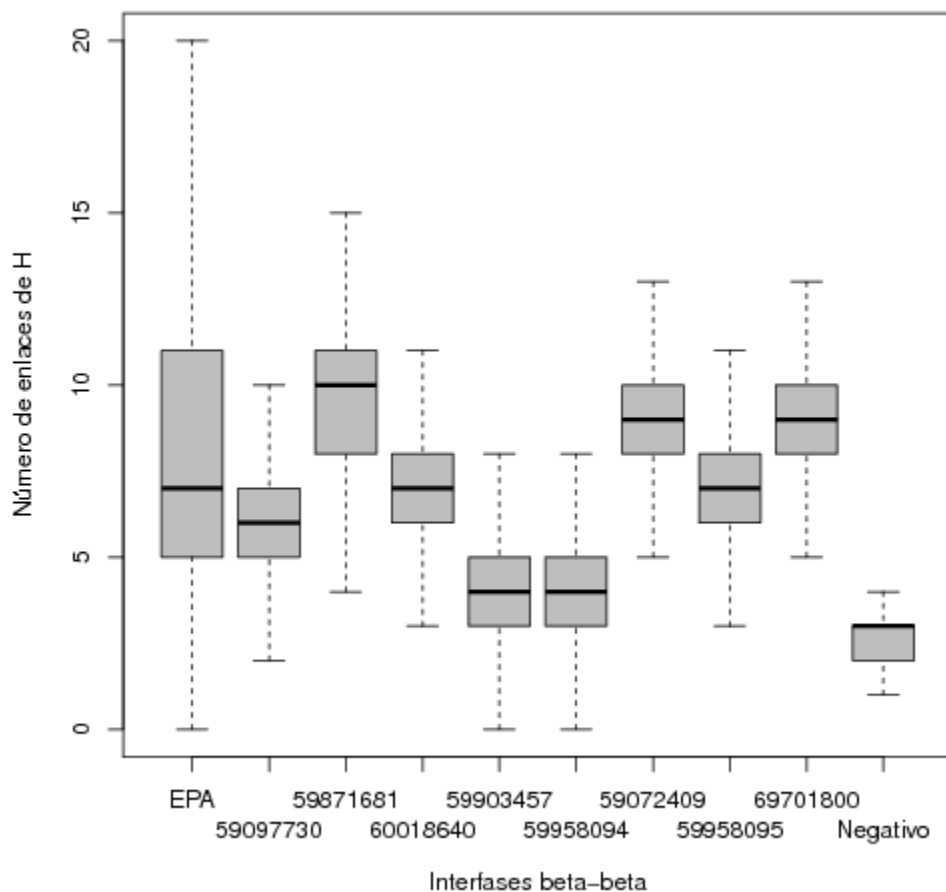


Figura 6.16. Diagrama de cajas del número de puentes de hidrógeno formados cada 2 ps en los diferentes sistemas. El valor de la mediana se muestra como una barra horizontal en color negro, las cajas representan el intervalo del primer al tercer cuartil, y las líneas discontinuas muestran los valores mínimos y máximos.

VI.2.6.3. Estructuras secundarias

La asignación de las estructuras secundarias por aminoácido a través del tiempo se graficaron para las cadenas B (Figura 6.17) y las cadenas D (Figura 6.18) de los diez sistemas homodiméricos empleando la metodología de asignación de estructura secundaria DSSP.^{55,56}

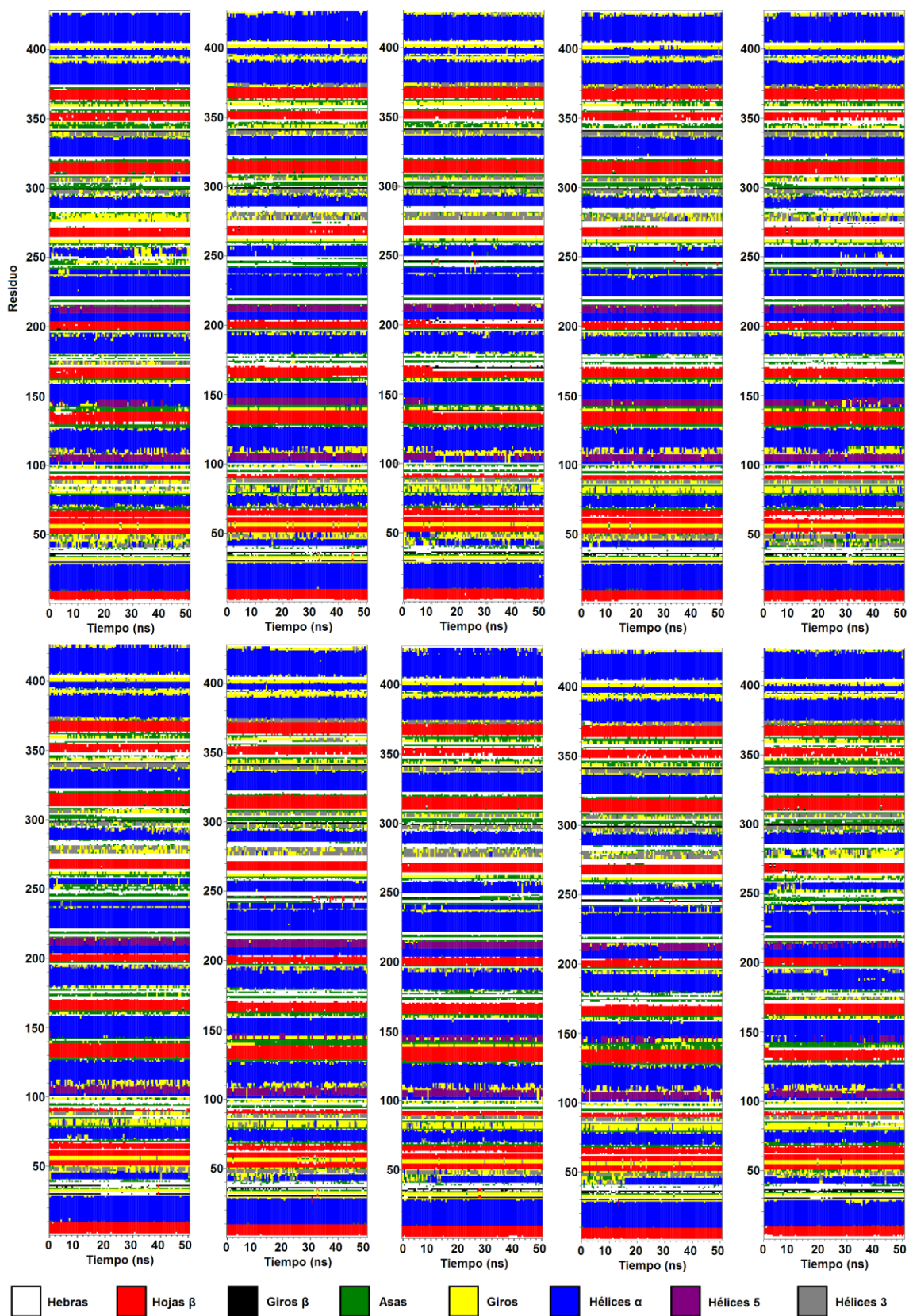


Figura 6.17. Asignación de estructuras secundarias a través del tiempo para las cadenas B.

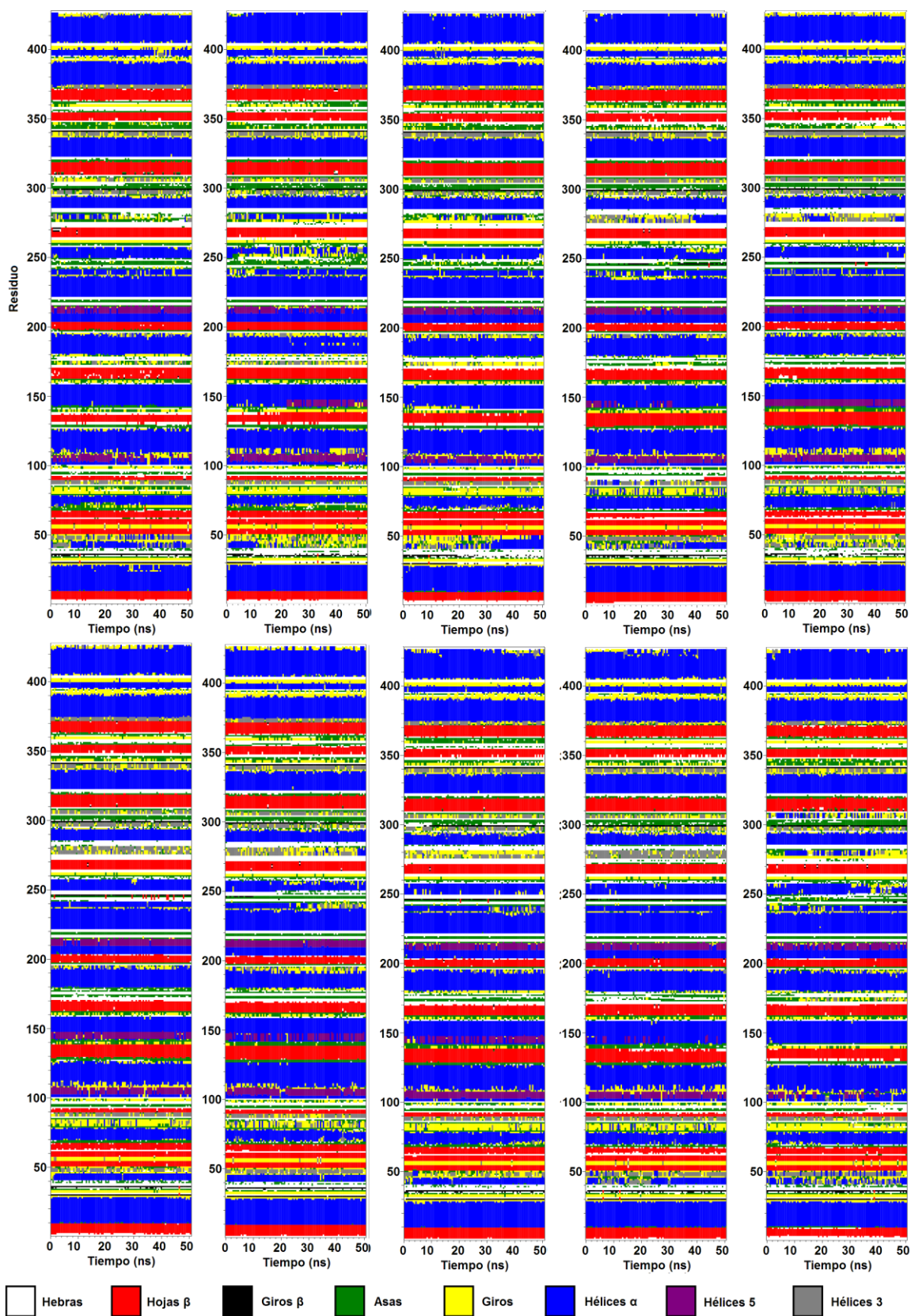


Figura 6.18. Asignación de estructuras secundarias a través del tiempo para las cadenas D.

VI.2.6.4. Cálculos de energía de unión (ΔH_{bind} y ΔG_{bind})

(a) Método MM/PBSA. Se calcularon las entalpías de unión de los ligandos en cada cadena respecto al sistema proteico mediante la aproximación lineal de la ecuación de Poisson-Boltzmann. Las contribuciones polares, apolares y de mecánica molecular se muestran en la Figura 12.8 del Anexo III.

(b) Método LIE. Se calculó el valor obtenido de la energía libre de unión del ligando $\Delta^2 G_{\text{bind}}$ por el método de expansión en series de LIE.

(c) Cálculo de energía mediante MM/PBSA y LIE. Para EPA y los ligandos con CID 60018640, 59072409, 59958095 y 69701800 se realizó la comparación entre los valores calculados. En la figura 6.19 se presenta la comparación de los dos métodos para el ligando en cada cadena, y la dispersión de los valores de energía de unión para estos ligandos (Figura 6.20).

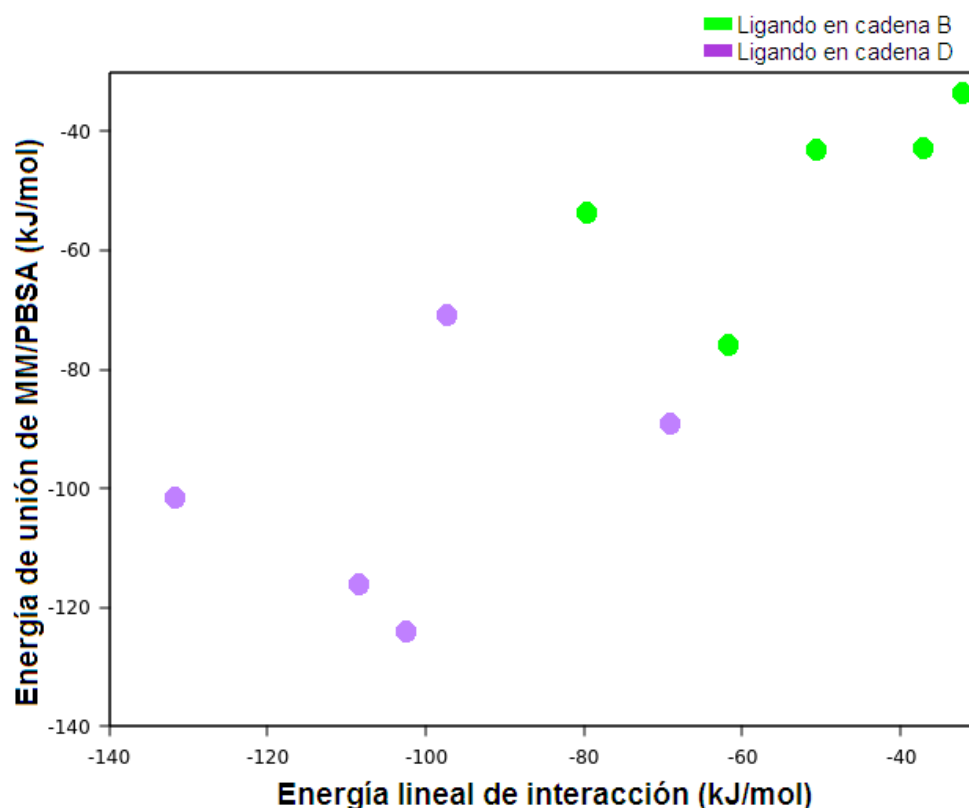


Figura 6.19. Comparación de los métodos de cálculo de energía MM/PBSA y LIE con un diagrama de dispersión.

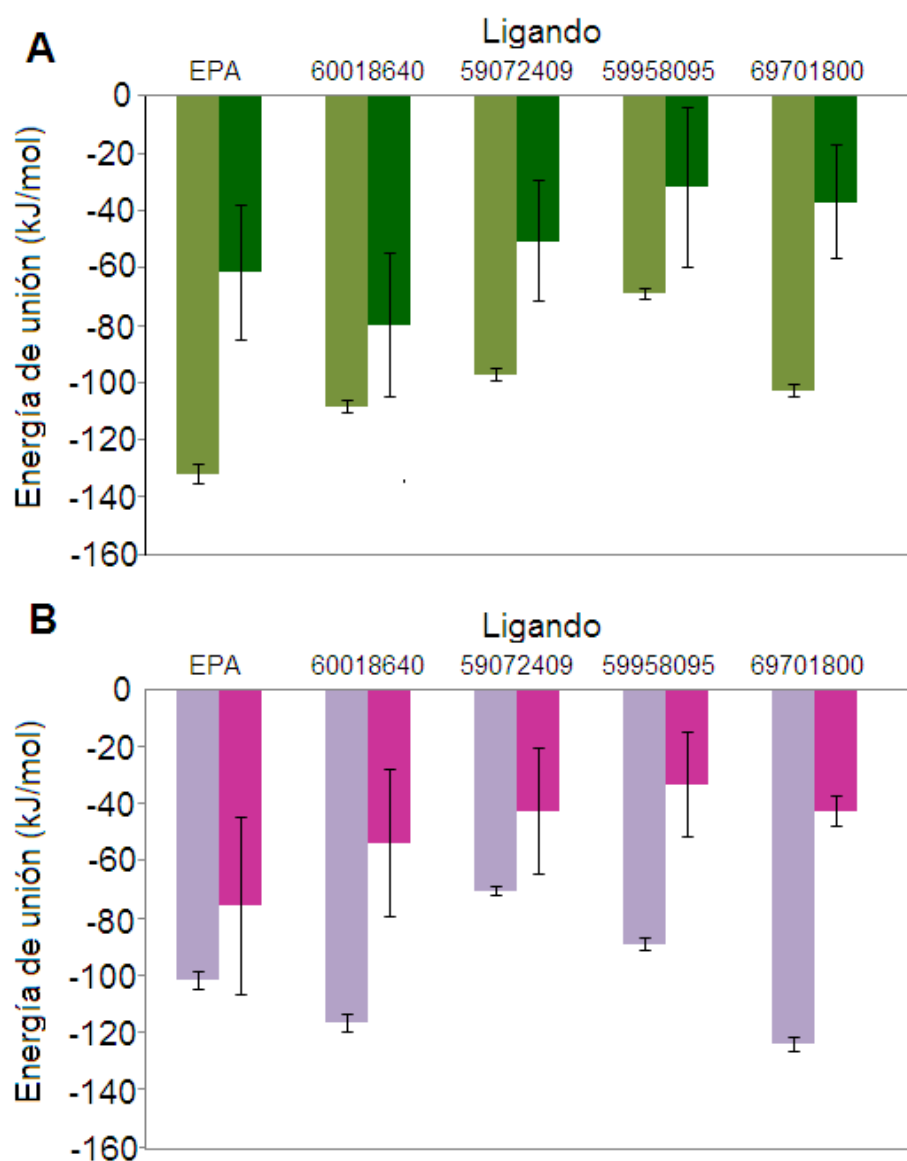


Figura 6.20. Comparación pareada de los métodos de cálculo de energía MM/PBSA (barra izquierda) y LIE (barra derecha). (A). Ligandos en la cadena B, en la interfase del homodímero β_2 . (B) Ligandos en la cadena D, expuestos parcialmente al solvente. Las barras indican la desviación estándar del cálculo.

VII. DISCUSIÓN

VII. 1. Primera parte

VII. 1. 1. Permeabilidad, hidrofobicidad y BHE

Para que un fármaco pueda ser absorbido de la circulación a un tejido, este debe estar disuelto. La permeabilidad del tejido en el sitio de absorción determina la cantidad del fármaco que lo atraviesa por algún tipo de transporte. En la barrera hematoencefálica (BHE), las células endoteliales de los vasos que irrigan el sistema nervioso central presentan uniones estrechas que dificultan el transporte paracelular y una alta expresión de complejos ABC que devuelven a la circulación los solutos que hayan ingresado al tejido por transporte pasivo.⁵⁷ Estas bombas de eflujo tienen afinidad por un gran número de compuestos endógenos y exógenos.⁵⁸ Las especies moleculares iónicas son las que presentan mayor dificultad para atravesar las membranas plasmáticas, si no ingresan mediante un transportador, debido al entorno hidrofóbico del interior de la bicapa lipídica. Los compuestos hidrofóbicos tienen mayor oportunidad de permear las células endoteliales de la BHE mediante mecanismos de difusión según la primera ley de Fick.⁵⁹ Si la estructura de estos le da características con alta lipofilicidad, los compuestos tienden a permanecer preferentemente en la región intermembranal que en los medios acuosos intracelular y extracelular, debido al efecto hidrofóbico. Dos medidas de la hidrofobicidad de una estructura son el logaritmo de la constante de reparto octanol-agua (logP) y el área superficial polar (PSA o tPSA).

El modelo lineal de Clark predice el transporte de un permeante a través de la BHE³² por difusión simple. En este trabajo se propusieron modelos lineales para la predicción del transporte utilizando el logP calculado mediante un método de fragmentos, el XLOGP3^{60,61} debido a las diferencias en los métodos de medición y el cálculo del descriptor logP que se encontraron.⁶⁰ Se estudió un grupo de 202 estructuras generado a partir de un conjunto inicial de 232 compuestos³³, del cual se generaron los grupos de estimación o entrenamiento y predicción o prueba con una división del 80% para el primer conjunto mediante una división empírica de los datos. Esto último como resultado de analizar los resultados preliminares de los modelos generados aleatoriamente al considerarse divisiones de 90%, 85%, 80%, 75%, 70%, 65% y 60%. De los descriptores moleculares calculados se generaron modelos lineales preliminares para escoger aquellos con mayor significancia estadística global y por regresor. De este análisis se seleccionaron dos modelos (ver Tablas 6.1, 6.2 y 6.3).

El modelo lineal de dos descriptores generado incluye logP y tPSA, de manera análoga al modelo de Clark, y el modelo de tres descriptores incluye además el descriptor tamaño máximo anular (*MaxSizeRing*) de variable discreta, el cual se generó mediante un análisis de regresión lineal paramétrico, que es permitido cuando existe un buen ajuste.⁶² Estos modelos lineales tuvieron mayor significancia al eliminar los valores atípicos por alguno de los tres métodos ensayados. Las

densidades de residuales estandarizados y estudentizados pasaron la prueba de normalidad, supuesto indispensable para el método,⁶³ y no mostraron diferencia estadística entre ellas. El método de distancias de Cook (D_i) no se utilizó debido a que no hay consenso en el cálculo de sus valores críticos y su densidad de probabilidad.⁶⁴ El modelo lineal con dos descriptores tiene las ventajas, respecto al modelo de Clark, en que tiene un error menor (0.261 *versus* 0.354³²), un mayor estadístico F_c (235.6 *versus* 95.8³²) y los valores de error estándar de los regresores también son menores.

El estadístico de validación interna q_{int}^2 es mayor a 0.5 en los seis casos, como es sugerido.^{36,65} Los cocientes de estadísticos de determinación son todos menores a uno y las pendientes k' de los respectivos modelos ajustados a cero de valores observados contra valores predichos están en el intervalo [0.85,1.15], por lo que los modelos QSPR se consideran predictivos.^{36,66}

Los modelos con dos y tres descriptores aplicados a los grupos de estructuras similares a EPA generaron dos grupos casi idénticos de estructuras (Figura 6.8), por lo cual se decidió mantener a lo largo del estudio el modelo más simple, el de dos descriptores.

VII. 1. 2. Relevancia de los descriptores logP y tPSA

La representación topológica de las estructuras moleculares mediante descriptores bidimensionales presenta las ventajas de la simplicidad de la información, la invariabilidad a la rototraslación molecular y la ausencia de la necesidad de optimización geométrica en un espacio no métrico.⁶⁷ Los descriptores moleculares 2D tienen un conjunto de características que les permiten ser usados ampliamente. De los descriptores que pueden calcularse por varios algoritmos está el logaritmo de la constante de reparto logP, que es un indicador de la hidrofobicidad mediante la solubilidad no acuosa del compuesto; y el área superficial polar topológica tPSA, que suma las áreas calculadas de N y O. Los valores bajos de tPSA están relacionados con la baja polaridad de la estructura. Estos descriptores representan adecuadamente las características de hidrofobicidad y polaridad de las que dependen los ligandos para atravesar la BHE.³²

VII. 1. 3. Predicción del transporte a través de la BHE

El análisis estadístico de los modelos QSPR mostró que los modelos con dos descriptores presentaban la mayor confiabilidad en las predicciones, por lo que con ellos se lograron filtrar 97 estructuras de los 5,163 compuestos con alta similitud con EPA de la base de datos de *PubChem*.

VII. 2. Segunda parte

VII. 2.1. Modelado de proteínas por homología

El modelado por homología es un método útil cuando se utilizan estructuras resueltas con alta similitud con la estructura primaria a modelar, ya que se usa la estructura resuelta como molde. La validación del modelado por homología de TUBA1A y TUBB2A incluye dos análisis empíricos, los valores z_c y el diagrama de Ramachandran. Los valores de los estadísticos z_c y z_c obtenidos de QMEAN se encuentran en el intervalo de modelados adecuados⁶⁸. Los valores de z_c de QMEAN se consideran aceptables cuando son lo más cercanos a cero, que se relacionan con un valor mayor al 80% de similitud, considerando la suma de porcentajes de residuos cuyas distancias entre $C\alpha$ son menores o iguales a 1, 2, 4 y 8 Å entre ambas estructuras.⁶⁹ El diagrama de Ramachandran (Figura 6.3) para la α -tubulina muestra el 96.6% de los residuos de aminoácidos en las regiones más favorecidas, en color rojo: A, B, L (370 residuos) y el 3.4% restante en las regiones favorecidas, en color amarillo: a, b, l, p. Para la β -tubulina el 95.4% de los residuos están en las regiones más favorecidas y el 4.6% en las regiones favorecidas. Las regiones más favorecidas de ambos modelos, los residuos de Gly están presentes en los giros beta, hebras sin estructura secundaria, algunos en hojas beta y en las flexiones de hélices (en inglés: *kinks*), que causan pérdida de helicidad. Los residuos de Pro se localizan principalmente en giros abruptos de asas, y en los extremos de las hélices. Las regiones no permitidas están caracterizadas por una energía potencial de torsión alta debida a conformaciones eclipsadas de los átomos alrededor de los enlaces que definen los ángulos diedros, repulsiones descritas por la energía de Lennard-Jones y repulsiones electrostáticas debidas a la cercanía de los átomos.

VII. 2.2. Caracterización dinámica del dímero $\alpha\beta$

El heterodímero $\alpha\beta$ es la unidad longitudinal mínima del MT (Figura 2.3). El sitio de unión de los taxenos (Figura 6.4) se encuentra expuesto al agua. Los valores del gráfico de RMSD de la cadena principal a través del tiempo (Figura 6.5A) muestra una tendencia creciente, indicando que no se alcanzó un equilibrio dinámico en el tiempo de simulación. El GDP permaneció en la cavidad N de la cadena β y el complejo $Mg^{II}[GTP]$ se alejó del sitio N en la cadena α , solvatándose en el medio acuoso, lo que sugiere que $\Delta G_{\text{solv}} < \Delta G_{\text{complejo}}$ para este sistema. Las fluctuaciones de las cadenas laterales muestran valores altos en múltiples regiones porque el sistema no alcanzó el equilibrio dinámico. Los extremos amino y carboxilo terminales presentan valores muy altos de RMSF debido a que están rodeados del solvente y no son estabilizados por otras regiones de la proteína. La integridad globular y la estructura terciaria se conservó durante la simulación, como se observa en el gráfico de radio de giro (Figura 6.5D), el cual se espera que no supere la mitad de la distancia de

separación máxima del dímero, cercano a 5.1 Å aproximando por simplicidad el centro de masa en el centro de una esfera (Figura 7.1).

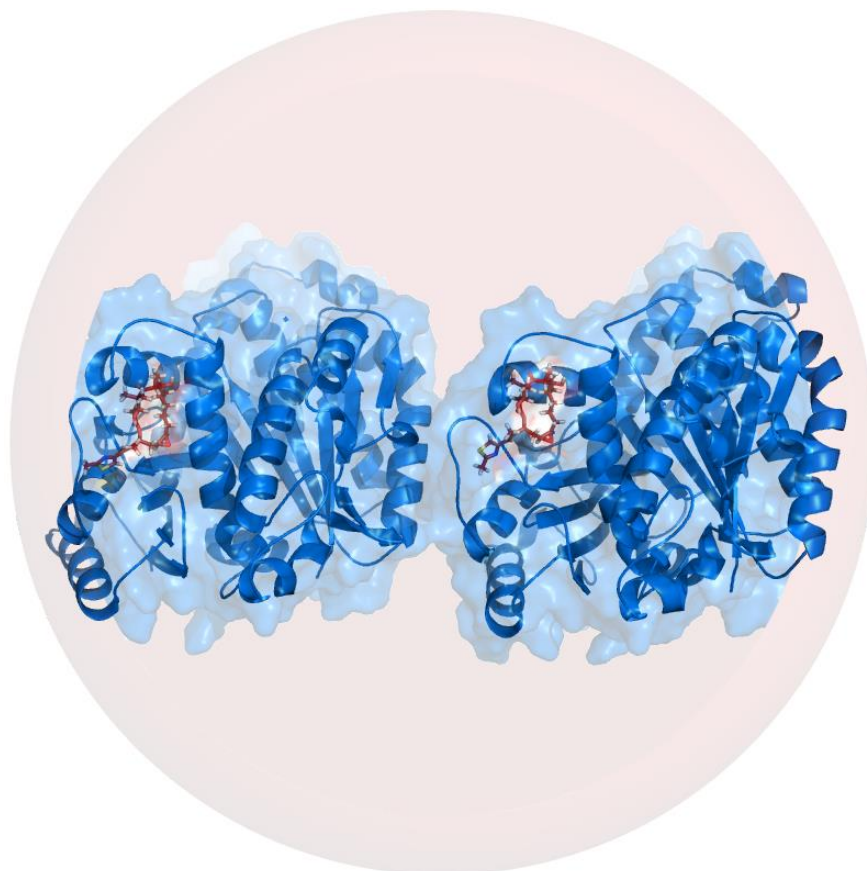


Figura 7.1. Modelo del homodímero β_2 representado en estructura terciaria y superficie de van der Waals en color azul. La estructura de EPA se muestra en barras color rojo. El radio de giro del dímero forma la esfera color rosa.

VII. 2. 3. Caracterización dinámica del heterotetrámero $\alpha_2\beta_2$

La inestabilidad de los sistemas heterodiméricos longitudinales en las simulaciones, debida de manera importante al área polar accesible al solvente (304.264 \AA^2) en el sistema $\alpha\beta$,⁷⁰ justificó una reducción de esta superficie y la estabilización del dímero a través de las interfases. Al construir un heterotetrámero $\alpha_2\beta_2$, los ligandos estabilizadores se encontrarían en una cavidad formada entre subunidades y no accesibles al agua (Figura 6.6). El área superficial polar accesible al solvente para la subunidad β_1 (cadena B, 180.495 \AA^2) y para la subunidad β_2 (cadena D, 192.840 \AA^2) disminuyó respecto al heterodímero longitudinal. El sistema heterotetramérico se construyó con EPA y sin ligando estabilizador. El sitio de unión de EPA en la cadena B está en la interfase con la cadena D y

poco accesible al agua. Este sistema heterotetramérico mostró mayor estabilidad que los sistemas heterodiméricos previos, por lo que se consideró el fragmento microtubular mínimo necesario para las simulaciones posteriores.

VII.2.3.1. Heterotetrámero $\text{GDP}_2, \text{GTP}_2\text{-}\alpha_2\beta_2$

El sistema $\alpha_2\beta_2$ (Figura 6.7) tuvo una mejora considerable respecto al sistema $\alpha\beta$ en cuanto a sus propiedades dinámicas. Los valores de la función RMSD disminuyeron aproximadamente a un cuarto de su magnitud respecto a los sistemas heterodiméricos, y muestran una tendencia al equilibrio dinámico desde los 10 ns de simulación. Los cofactores nucleotídicos tuvieron dos comportamientos, $\text{Mg}^{II}[\text{GTP}]$ en las cadenas A y C se mantuvieron en los sitios N con menores desviaciones que $\text{Mg}^{II}[\text{GDP}]$. El $\text{Mg}^{II}[\text{GTP}]$ tiene mayores interacciones con los aminoácidos D98, T145 y S140, adyacentes al fosfato γ y el fosfato β respectivamente. Además, las subunidades β restringen estéricamente la solvatación del complejo nucleotídico. Las fluctuaciones de las cadenas laterales muestran tres regiones de alta movilidad, correspondientes a hebras sin estructura secundaria definida y en contacto con el solvente. Las estructuras terciarias definidas no se desestabilizaron, y algunas asas desordenadas adquirieron estructuras secundarias en intervalos cortos de tiempo. El tamaño de la cavidad de unión de EPA, que varía a través del tiempo, fue un criterio para diseñar los acoplamientos moleculares posteriores. La cavidad de unión de EPA, con más de 0.97 nm de distancia entre los centros de masa de los aminoácidos adyacentes, donde EPA puede acceder sin impedimento estérico, se presentó de manera recurrente hacia la mitad de la simulación.

VII.2.3.2. Heterotetrámero $\text{EPA}_2\text{-GDP}_2, \text{GTP}_2\text{-}\alpha_2\beta_2$

El sistema $\text{EPA}_2\text{-}\alpha_2\beta_2$, en presencia del ligando en el sitio de la estructura resuelta en PDB: 4O4L, mostró una mayor estabilidad comparada con el sistema sin el ligando estabilizador. La cadena principal alcanzó un equilibrio dinámico, y los ligandos permanecieron en el sitio de unión; la estructura ubicada en la cadena B tuvo mayor estabilidad que la estructura de la cadena D, debido al aumento de las interacciones en la interfase β - β . Al igual que el heterotetrámero sin agente estabilizador, las estructuras secundarias y terciarias definidas se conservaron a lo largo de la simulación.

VII. 2. 4. Agrupamiento (*clustering*) y acoplamiento molecular de compuestos similares a EPA

Del agrupamiento o *clustering* (Tabla 6.5) se seleccionaron los confórmers provenientes de la simulación que presentaban el sitio de unión de EPA en la conformación abierta (distancia mínima de 0.9 nm entre los centros de masa de Q279 y L361). Posteriormente, se empleó un criterio empírico de elección del rango o *cutoff* de 0.09 nm, de tal manera que el número de grupos no fuera muy grande ni muy pequeño, y que el número de conformaciones integrantes fuera disminuyendo con tendencia similar a la distribución exponencial. Esta selección dio como resultado un total de 31 grupos, de los cuales se extrajeron las estructuras más representativas para ser empleadas en el estudio de acoplamiento.

El acoplamiento de los ligandos en las conformaciones medias de la subunidad de β -tubulina se realizó con el programa AutoDock Vina, que aunque su SF es menos precisa que la de otros programas, requiere un tiempo computacional menor, considerando que se realizarían 3,007 procesos de acoplamiento (97 estructuras filtradas en 31 heterotetrámeros de tubulina). De manera óptima, el acoplamiento se realizaría en el sistema heterotetramérico completo, pero debido a su complejidad, la imprecisión del cálculo SF y de la conformación del ligando acoplado en la subunidad β se rectificó en las simulaciones de MD posteriores.

Por su parte, los valores de similitud de interacciones obtenidos a partir del análisis de PLIFs fueron menores a 0.6 en todos los casos, consistiendo principalmente de interacciones hidrofóbicas. La significancia estadística de los coeficientes de similitud, como aquellos calculados con el coeficiente de Tanimoto a partir de huellas digitales booleanas, depende del tamaño poblacional y muestral de estructuras,⁷¹ por lo que no se realizaron pruebas de hipótesis a los valores de similitud obtenidos.

VII. 2. 5. Suficiencia de β_2 para la simulación. Necesidad de controles en la simulación

El tiempo computacional invertido en el sistema heterotetramérico es alto, aproximadamente 2.5 ns por día. La interfase β - β crea un entorno adecuado para la estabilidad del ligando unido a la cadena B, inclusive el ligando de la cadena D expuesto al solvente también participa en la estabilización por un fenómeno similar a la cooperatividad. Para evaluar los ocho compuestos con mejores valores de PLIFs y SF en simulación de MD, se redujo el sistema heterotetramérico al homodímero β_2 , o dímero transversal, con la finalidad de disminuir el costo computacional. El homodímero posee las características suficientes para la estabilización del complejo, y en particular para el acoplamiento del ligando, por lo que fue utilizado únicamente para evaluar al estabilizador. Esta inferencia inductiva implicó la necesidad de un control positivo, un homodímero con EPA, y un control negativo; para verificar la comparabilidad de $\alpha_2\beta_2$ y β_2 .

VII.2.5.1. Condiciones del equilibrio dinámico

El equilibrio dinámico se puede aproximar visualmente en los gráficos de la función RMSD(t) de las cadenas principales. Cuando las cadenas principales se encuentran en equilibrio, se definen los intervalos de tiempo donde se analizan las funciones RMSF(t) y los cálculos de energía libre de unión de los ligandos. El tiempo de equilibrio no coincide con una etapa de la simulación, sino que es particular para cada sistema. La aplicación de la DFT a las sucesiones de RMSD de las cadenas principales genera un gráfico suavizado, donde es más evidente el equilibrio dinámico. Los gráficos de DFT generan una función periódica, en la cual se busca el intervalo donde la amplitud de pico a pico es aproximadamente equidistante de una función constante, una línea horizontal en el gráfico. Los sistemas homodiméricos alcanzaron el equilibrio dinámico durante el tiempo de simulación, y dos de ellos, los sistemas con los ligandos con CID: 59903457 y CID: 59958094 (Figuras 6.12E y 6.12H) las cadenas B tuvieron incremento en la desviación cuadrática media. El control positivo, el sistema EPA₂-β₂ (Figura 6.12A) presentó una estabilidad equiparable a la del sistema heterotetramérico EPA₂-α₂β₂, y de manera análoga la estructura de EPA interfásial tuvo menor desviación respecto a la estructura expuesta al solvente. El control negativo, el sistema β₂ (Figura 6.12J) es también estable, con valores de RMSD de equilibrio alrededor de 0.2 nm y gráficos RMSF similares (Figura 6.14A y J). La diferencia notable entre los dos sistemas es la compactación de los dímeros, el control positivo tuvo valores de radio de giro menores a los del control negativo, que indica mayor número de interacciones intermoleculares e intramoleculares (Figura 6.15A y J).

VII.2.5.2. Puentes de hidrógeno. Comparación de sistemas con EPA y con control negativo

Los valores obtenidos de radio de giro del control positivo y de los ocho sistemas con ligandos son menores al control negativo debido a la formación de puentes de hidrógeno en la interfase β-β (Figura 6.16). El control positivo es el sistema que puede formar una mayor cantidad de puentes de hidrógeno interfásiales durante el tiempo de simulación, hasta veinte puentes de hidrógeno predichos por las distancias interatómicas. Se compararon las distribuciones discretas de los enlaces de hidrógeno formados en los sistemas con los ocho ligandos evaluados a través de sus medianas. La formación de enlaces de hidrógeno intermoleculares encontrados en el ligando con CID: 60016640 es equiparable a los encontrados para la EPA. Los puentes de hidrógeno establecidos por los ligandos con CID: 59871681, 59072409 y 69701800 tuvieron diferencias en la comparación de sus medianas con el control positivo, siendo mayores que la mediana de los puentes de hidrógeno de éste. El control positivo y los ocho sistemas evaluados son superiores al control negativo, siendo la mediana de sus puentes de hidrógeno menor en todos los casos. La formación de un gran número de puentes de hidrógeno estabiliza al dímero β₂, requiriendo una perturbación mayor para disociarlo que la

necesaria para desestabilizar un dímero con pocos enlaces intermoleculares, como el control negativo.

VII.2.5.3. Estructuras secundarias a través del tiempo

Los diagramas de estructuras secundarias a través del tiempo (Figura 6.17) muestran que éstas se mantienen a través del tiempo de simulación. En las cadenas B y D del homodímero, las regiones de residuos 100 a 110 y 140 a 150 presentaron variabilidad en su estructura secundaria. Estos segmentos se encuentran en la superficie externa del correspondiente MT, y en la estructura PDB: 4O4L están cercanas o en contacto a la MAP estatmina 1 que estabiliza las regiones, en particular los segmentos hidrofóbicos con varios residuos de G y uno de W. La hélice torcida en el segmento 100 a 110 está presente en todos los sistemas, excepto en el control negativo en la cadena D, y se desestabiliza en la cadena B del sistema con ligando CID: 59871681. El segmento helicoidal de 250 a 260 mostró inestabilidad en ambas cadenas. Esta hélice, con menos de dos giros, tiene contacto con la α -tubulina en sistemas heteromultiméricos, y es estabilizada con los residuos hidrofóbicos A403, W407 y V409. El segmento 300 a 305 se encuentra en la interfase β - β , por lo que en la cadena D se espera la variación de la estructura al encontrarse sus residuos hidrofóbicos en contacto con el solvente. El segmento 390 a 400, con los residuos A393, F394 y L395 presenta inestabilidad porque en sistemas heteromultiméricos está en contacto con la α -tubulina. El segmento 270-286, el asa M, no tuvo tendencias particulares entre los sistemas, pero el entorno hidrofóbico creado por la interfase β - β mostró ser importante en la estabilización de los sistemas de tubulinas.

Las variaciones en la estructura secundaria se deben fundamentalmente al modelo del sistema, que simplifica los complejos macromoleculares de los MT donde están presentes interfases estabilizadoras múltiples.

VII.2.6. Energías de unión. Comparación entre aproximaciones MM/PBSA y LIE

El cálculo de energía libre implica invertir una gran cantidad de tiempo computacional, por lo que la aproximación de Debye-Hückel de la ecuación de Poisson-Boltzmann para la energía de solvatación considera fuerza iónica baja, sustituyendo el término exponencial por uno lineal, aunque para sistemas con cargas altas se requiere la resolución de la ecuación diferencial parcial no lineal de PB.⁵⁰ El método MM/PBSA realiza un cálculo para el cambio de entalpía de unión (ΔH_{bind}), y se considera el término entrópico como despreciable, porque el cálculo requiere recursos computacionales elevados.

La principal diferencia entre los métodos MM/PBSA y LIE es que los términos electrostáticos en el segundo se obtienen de las constantes dieléctricas a través de la ecuación PB.⁷² La comparación de los valores energéticos no son equivalentes estadísticamente ($p \sim 10^{-25}$). Las cadenas B y D

alcanzaron el equilibrio dinámico en distintos intervalos de tiempo, los cálculos MMPBSA y LIE se realizaron en el intervalo de tiempo de 45 ns a 50 ns, donde las cadenas B y D tenían comportamientos distintos en sus desviaciones medias, detectados con los gráficos con DFT aplicada. El equilibrio dinámico influye en el cálculo energético, por lo que las diferentes contribuciones de los métodos no mostraron ser equivalentes en los dos subsistemas. Se considera el método LIE una aproximación que pierde información al incluir únicamente un término lineal para estimar Δ^2G_{bind} , además de que el término entrópico en MMPBSA se omitió y sólo se calculó ΔH_{bind} . La desviación estándar de los cálculos con el método LIE es mayor que con el método MM/PBSA, y en los casos analizados los valores de energía de unión por el método LIE son menores a los valores obtenidos con el método MM/PBSA. La correlación entre los valores por ambos métodos no es lineal.

VII. 2. 7. Elección del *hit* computacional

VII.2.7.1. Relaciones estructurales

Los ligandos con CID: 59097730, 59072409, 59958094 y 69701800 tuvieron los menores valores de SF en el acoplamiento, y en las simulaciones de MD estabilizan los sistemas β_2 . Además los ligandos 59072409 y 59958094 mantuvieron a las cadenas laterales de sus homodímeros con fluctuaciones medias (RMSF) menores durante el equilibrio, y conservaron la estructura homodimérica con un radio de giro de aproximadamente 3.3 nm, lo que indica que la proteína no experimentó desviaciones importantes respecto a su centro de masa durante la simulación MD. También el control positivo y el sistema con el ligando 59097730 alcanzaron una estabilidad conformacional, aunque con mayores perturbaciones hacia la mitad de la simulación MD.

El sistema en presencia del ligando CID: 59871681 tuvo la mayor capacidad de formar puentes de hidrógeno (alrededor del tiempo de 21 ns) entre los monómeros β , aunque es estadísticamente equivalente a EPA y para el sistema en presencia del ligando con CID: 59903457 se encontraron los valores más bajos de cambio de entalpía de unión, en la conformación tomada a los 27 ns, para ambas cadenas.

Los ligandos con alta similitud a EPA, filtrados con el modelo lineal de dos descriptores, comparten varias subestructuras, como el 2-metiltiazol, un macrólido de 16 a 18 miembros, y enlaces dobles (*Z*) y (*E*). Los PLIFs y los farmacóforos propuestos por Pharmit⁵¹ coinciden en la asignación del 2-metiltiazol, el metileno 2, el grupo metilo en posición $\omega-1$, el grupo carbonilo o hidroxilo en posición 3 o 5, y los grupos metilo en posiciones 5-6 a 8. Los grupos metilo y metileno interactúan con residuos hidrofóbicos del sitio, principalmente Leu, Pro y metilenos β a γ de Gln. Los grupos carbonilo que interactúan con hidrógenos polares y el átomo de nitrógeno del tiazol, al tener cierto carácter nucleofílico debido a que el par electrónico no participa en el sistema aromático, forma puentes de hidrógeno con la T274. Para la cavidad de unión de EPA en las conformaciones con distancia mayor

a 0.9 nm, los ligandos deben poseer los subespacios farmacofóricos hidrofóbicos y al menos un grupo capaz de formar puentes de hidrógeno con T274 y Q279.

Los ligandos con mejor comportamiento estabilizador son en primer lugar: 69701800, en segundo lugar 59072409 y en tercer lugar 59097730.

VII.2.7.2. Reacciones de biotransformación

Es improbable que la EPA llegue al sitio de unión en β -tubulina como epóxido⁵² debido a la acción de las enzimas de biotransformación. De manera análoga, el ligando con CID: 59097730 (Figura 7.2A) puede ser sustrato de enzimas de biotransformación como esterasas y citocromos. Las reacciones de fase I predichas incluyen la hidrólisis de lactona (Figura 7.2B) y la oxidación del enlace doble 10-*trans* (Figura 7.2C), formando un *vec*-diol. Por sus características predominantemente hidrofóbicas también se espera que este ligando sea sustrato de enzimas de fase II. La inactivación de los ligandos por esterasas podría evitarse con la sustitución de la lactona por una lactama, de manera similar a la ixabepilona, que es el isómero clásico de la epotilona B. La lactona también puede sustituirse por bioisómeros como sultonas, sultamas, triazoles y otros grupos. La formación de epóxidos puede dificultarse si se sustituyen los enlaces dobles *cis* por alcoholes o *vec*-dioles. El grupo trifluorometilo, al ser un isómero no clásico del metilo, tiene la ventaja de modificar la probabilidad de oxidación del enlace doble (*Z*), porque aunque tiene un radio de van der Waals similar, tiene efectos electrónicos diferentes.⁷³ El valor NOR (del inglés: *Normalised Occurrence Ratio*) es la probabilidad condicional normalizada de que se dé la reacción una vez que se ha llegado al sitio de biotransformación. El valor NOR de que reaccione el enlace doble con grupo trifluorometilo es 0.0,⁵² por lo que la formación de un epóxido en este grupo es muy improbable.

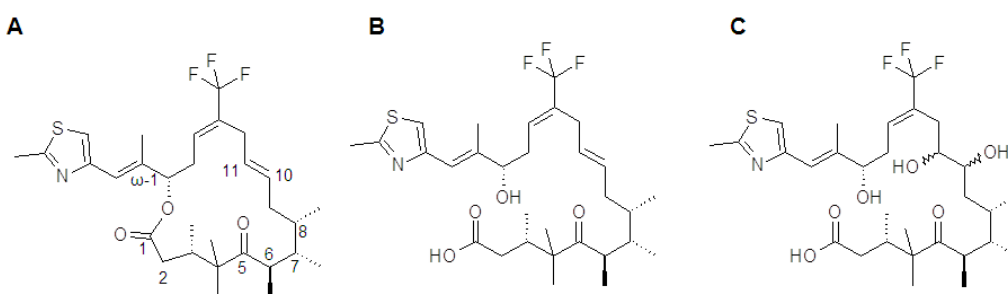


Figura 7.2. Predicción de las reacciones de biotransformación⁵² para el ligando con CID: 59097730 (A). Hidrólisis de la lactona (B) y formación de un *vec*-diol (C).

El ligando con CID: 59871681 puede hidrolizarse y desactivarse definitivamente (NOR>0.6), por lo que no tiene potencial como estabilizador. Por otra parte, el ligando con CID: 60018640 también puede experimentar hidrólisis del éster y del carbonato, generando 2,2,2-tricloroetanol e hidrato de cloral, productos con acción sedante en el SNC y con alta hepatotoxicidad. De manera análoga, los

enlaces dobles pueden formar epóxidos. Este compuesto sería rápidamente degradado y con toxicidad alta.

El compuesto CID: 59903457 tiene baja solubilidad acuosa, y los grupos laurato y trideciloato se hidrolizarían ante esterasas, generando un epímero de la epotilona D. Además de la hidrólisis del éster, el enlace doble del ligando con CID: 59958094 puede hidratarse, siguiendo la regla de Markovnikov.⁷⁴ El ligando con CID: 59072409 (Figura 7.3A) también puede ser hidrolizado por esterasas (Figura 7.3B). Las reacciones de fase II (Figura 7.3C-F) pueden ocurrir con mayor probabilidad al hidrolizarse el ligando, debido al menor impedimento estérico en el 3-hidroxilo. Las reacciones de fase II permiten que ligandos hidrofóbicos adquieran mayor solubilidad acuosa para su excreción.

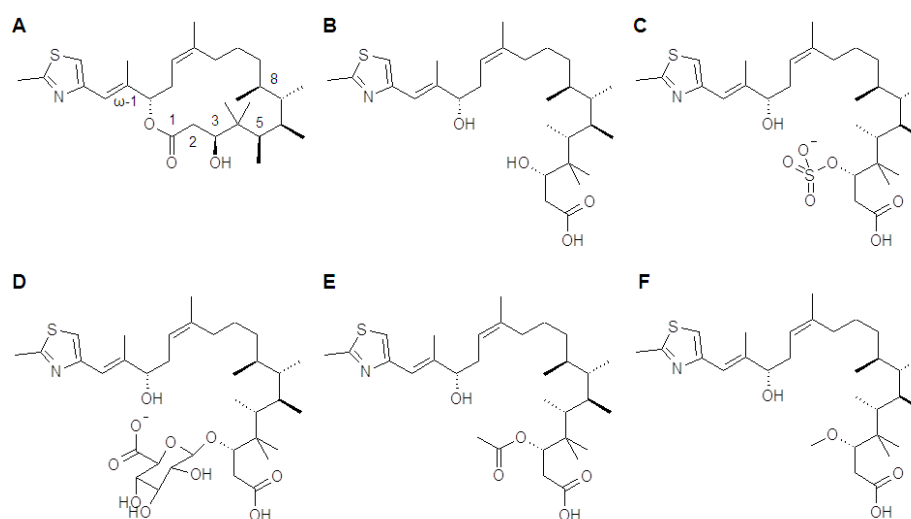


Figura 7.3. Predicción de las reacciones de biotransformación⁵² para el ligando con CID: 59074402 (A), hidrólisis de la lactona (B) y reacciones de fase II de biotransformación como sulfatación (C), glucuronidación (D), acetilación (E). También se espera al menos una metilación (F).

El ligando con CID: 59958095 (Figura 7.4) tiene una alta probabilidad de que su enlace doble *cis* forme un epóxido, producto electrofílico que reaccionaría espontáneamente con casi cualquier nucleófilo formando un enlace covalente e inactivándose mutuamente.

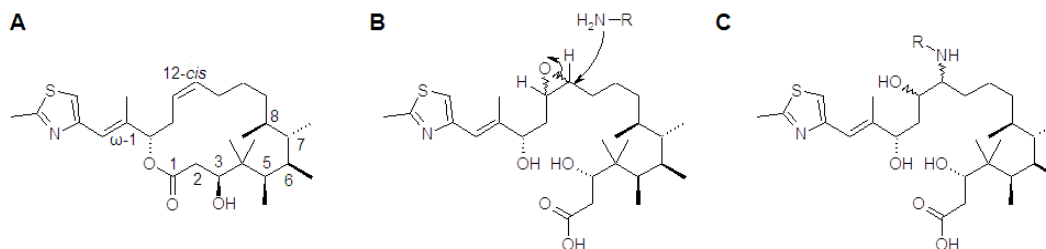


Figura 7.4. Predicción de las reacciones de biotransformación⁵² para el ligando con CID: 59958095 (A). El enlace doble *cis* se oxida fácilmente a un epóxido (B) que reacciona con nucleófilos (C).

El ligando con CID: 69701800 (Figura 7.5) también puede ser sustrato de esterasas. La sustitución de un enlace doble (*Z*) por el anillo de ciclopropano en la posición 12 del macrólido impide la generación de un epóxido o un *vec*-diol. La hidroxilación del 2-metilthiazol podría interferir con la actividad en este subespacio farmacofórico.

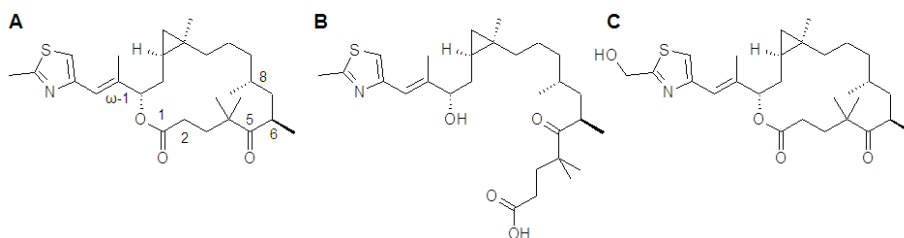


Figura 7.5. Predicción de las reacciones de biotransformación⁵² para el ligando con CID: 69701800 (A). Se espera la hidrólisis de la lactona (B) y la hidroxilación del 2-metilthiazol (C).

Por tanto, los ligandos acoplados con las mejores características estructurales potenciales para estabilizar los sistemas de tubulinas son 59097730 y 59074402 (Figuras 7.2A y 7.3A).

VII.2.6.3. Impacto de las interacciones de los ligandos con las isoformas de β -tubulina

Los residuos de aminoácidos implicados en las interacciones con β -tubulina (Tabla 6.7) descritos en los PLIFs están presentes en las isoformas I, IIa, IIb, III, IVa, IVb, V, VI y VIII (Figura 7.6), lo que sugiere una interacción probable de los ligandos similares a EPA. La isoforma I es de expresión ubicua, la V es de expresión cardíaca y la VI en el linaje hematopoyético,^{14,25} lo que estaría involucrado en posible cardiotoxicidad y mielotoxicidad. Por otra parte, se ha reportado que la isoforma III posee mayor rigidez conformacional que los otros tipos de β -tubulina.⁷⁵

Las interacciones proteína-ligando son en su mayoría de tipo hidrofóbico, donde sólo algunos de los ligandos involucran interacciones de puente de hidrógeno. La formación de un puente de hidrógeno con T274 es una interacción importante, y los residuos de interacción, con excepción de R276, están presentes en estas isoformas. Sólo el ligando con CID: 69701800 interactúa con R276, por lo que su cambio por Gln en β VI (TUBB1) podría definir la selectividad hacia esta isoforma de expresión hematopoyética.

P07437	TBB5_HUMAN	181	EPYNATLSVHQLVENTDETYCIDNEALYDICFRTLKLLTTPPTYGDLNHLVVSATMSGVTTCL	240
Q13885	TBB2A_HUMAN	181	EPYNATLSVHQLVENTDETYCIDNEALYDICFRTLKLLTTPPTYGDLNHLVVSATMSGVTTCL	240
Q9BVA1	TBB2B_HUMAN	181	EPYNATLSVHQLVENTDETYCIDNEALYDICFRTLKLLTTPPTYGDLNHLVVSATMSGVTTCL	240
Q13509	TBB3_HUMAN	181	EPYNATLSIHQLVENTDETYCIDNEALYDICFRTLKLLATPTYGDLNHLVVSATMSGVTTSL	240
P04350	TBB4A_HUMAN	181	EPYNATLSVHQLVENTDETYCIDNEALYDICFRTLKLLTTPPTYGDLNHLVVSATMSGVTTCL	240
P68371	TBB4B_HUMAN	181	EPYNATLSVHQLVENTDETYCIDNEALYDICFRTLKLLTTPPTYGDLNHLVVSATMSGVTTCL	240
Q9BUF5	TBB6_HUMAN	181	EPYNATLSVHQLVENTDETYCIDNEALYDICFRTLKLLTTPPTYGDLNHLVVSATMSGVTTSL	240
Q9H4B7	TBB1_HUMAN	181	EPYNAVLSIHQLIENADACFCIDNEALYDICFRTLKLLTTPPTYGDLNHLVSLTMSGITLST	240
Q3ZCM7	TBB8_HUMAN	181	EPYNATLSVHQLIENADETFCDNEALYDICKSKLKLPTTPPTYGDLNHLVVSATMSGVTTCL	240
			*****:****:*:* :*****:**** *****:****:****	
P07437	TBB5_HUMAN	241	RFPGQLNADLRKLAVMNVPFPRLLHFFMPGFAPLTSRGSQOYRALTVPELTQQVFDKAKNMM	300
Q13885	TBB2A_HUMAN	241	RFPGQLNADLRKLAVMNVPFPRLLHFFMPGFAPLTSRGSQOYRALTVPELTQQMFDKSNMM	300
Q9BVA1	TBB2B_HUMAN	241	RFPGQLNADLRKLAVMNVPFPRLLHFFMPGFAPLTSRGSQOYRALTVPELTQQMFDKSNMM	300
Q13509	TBB3_HUMAN	241	RFPGQLNADLRKLAVMNVPFPRLLHFFMPGFAPLTSRGSQOYRALTVPELTQQMFDKAKNMM	300
P04350	TBB4A_HUMAN	241	RFPGQLNADLRKLAVMNVPFPRLLHFFMPGFAPLTSRGSQOYRALTVPELTQQMFDKAKNMM	300
P68371	TBB4B_HUMAN	241	RFPGQLNADLRKLAVMNVPFPRLLHFFMPGFAPLTSRGSQOYRALTVPELTQQMFDKAKNMM	300
Q9BUF5	TBB6_HUMAN	241	RFPGQLNADLRKLAVMNVPFPRLLHFFMPGFAPLTSRGSQOYRALTVPELTQQMFDARNTMM	300
Q9H4B7	TBB1_HUMAN	241	RFPGQLNADLRKLAVMNVPFPRLLHFFMPGFAPLTAQGSQOYRALTVAELTQQMFDARNTM	300
Q3ZCM7	TBB8_HUMAN	241	RFPGQLNADLRKLAVMNVPFPRLLHFFMPGFAPLTSRGSQOYRALTVAELTQQMFDKAKNMM	300
			*****:****:*:* :*****:**** *****:****:****	
P07437	TBB5_HUMAN	301	AACDPRHGRYLTVAAVFRGRMSMKEVDEQMLNVQNKNSYFVEWIPNNVKTAVCDIPPRG	360
Q13885	TBB2A_HUMAN	301	AACDPRHGRYLTVAAIIFRGRMSMKEVDEQMLNVQNKNSYFVEWIPNNVKTAVCDIPPRG	360
Q9BVA1	TBB2B_HUMAN	301	AACDPRHGRYLTVAAIIFRGRMSMKEVDEQMLNVQNKNSYFVEWIPNNVKTAVCDIPPRG	360
Q13509	TBB3_HUMAN	301	AACDPRHGRYLTVAIVFRGRMSMKEVDEQMLAIQSKNSYFVEWIPNNVKVAVCDIPPRG	360
P04350	TBB4A_HUMAN	301	AACDPRHGRYLTVAAVFRGRMSMKEVDEQMLSVQSKNSYFVEWIPNNVKTAVCDIPPRG	360
P68371	TBB4B_HUMAN	301	AACDPRHGRYLTVAAVFRGRMSMKEVDEQMLNVQNKNSYFVEWIPNNVKTAVCDIPPRG	360
Q9BUF5	TBB6_HUMAN	301	AACDPRHGRYLTVAIVFRGRMSMKEVDEQMLAIQSKNSYFVEWIPNNVKVAVCDIPPRG	360
Q9H4B7	TBB1_HUMAN	301	AACDLRRGRYLTVAICIFRGMSTKEVDQQLSVQTRNSSFVEWIPNNVKVAVCDIPPRG	360
Q3ZCM7	TBB8_HUMAN	301	AACDPRHGRYLTAAAIIFRGRMPMREVDEQMFNIQDKNSYFADWLPNNVKTAVCDIPPRG	360
			**** *:*****:****:****:****:****:****:****:****:****:****:****	
P07437	TBB5_HUMAN	361	LKMAVTFIGNSTAIQELFKRISEQFTAMFRRKAFHLHWYTGEGMDEMEFTEAESNMNDLVS	420
Q13885	TBB2A_HUMAN	361	LKMSATFIGNSTAIQELFKRISEQFTAMFRRKAFHLHWYTGEGMDEMEFTEAESNMNDLVS	420
Q9BVA1	TBB2B_HUMAN	361	LKMSATFIGNSTAIQELFKRISEQFTAMFRRKAFHLHWYTGEGMDEMEFTEAESNMNDLVS	420
Q13509	TBB3_HUMAN	361	LKMSSTFIGNSTAIQELFKRISEQFTAMFRRKAFHLHWYTGEGMDEMEFTEAESNMNDLVS	420
P04350	TBB4A_HUMAN	361	LKMAATFIGNSTAIQELFKRISEQFTAMFRRKAFHLHWYTGEGMDEMEFTEAESNMNDLVS	420
P68371	TBB4B_HUMAN	361	LKMSATFIGNSTAIQELFKRISEQFTAMFRRKAFHLHWYTGEGMDEMEFTEAESNMNDLVS	420
Q9BUF5	TBB6_HUMAN	361	LKMASTFIGNSTAIQELFKRISEQFSAMFRRKAFHLHWYTGEGMDEMEFTEAESNMNDLVS	420
Q9H4B7	TBB1_HUMAN	361	LSMAATFIGNSTAIQEIFNRVSEHFSAMFRRKAFVHWYTGEGMDEMEFTEAESNMNDLVS	420
Q3ZCM7	TBB8_HUMAN	361	LKMSATFIGNSTAIQELFKRVSEQFTAMFRRKAFHLHWYTGEGMDEMEFTEAESNMNDLVS	420
			.:*****:****:****:****:****:****:****:****:****:****:****	

Figura 7.6. Alineamiento de las estructuras primarias de las isoformas de β -tubulina. Se muestran los códigos UniProt P07437: β I, Q13885: β IIa, Q9BVA1: β IIb, Q13509: β III, P04350: β IVa, P68371: β IVb, Q9H4B7: β VI, Q3ZCM7: β VIII.

Las interacciones a distancia, los potenciales termodinámicos y la actividad de MAPs definen de forma más verosímil las interacciones y afinidades de los ligandos con las diferentes isoformas de β -tubulina, no sólo la estructura primaria. Los cálculos de términos entrópicos y la presencia de MAPs en las simulaciones aproximarían en mayor medida la afinidad de los ligandos por una u otra isoforma, y reducirían las interacciones en sitios no deseados como la médula ósea y las mucosas.

Como resultado de este análisis de los ligandos, sus interacciones y sus propiedades, los hits computacionales obtenidos (Figura 7.7) son 69701800, 59072409 y 59097730.

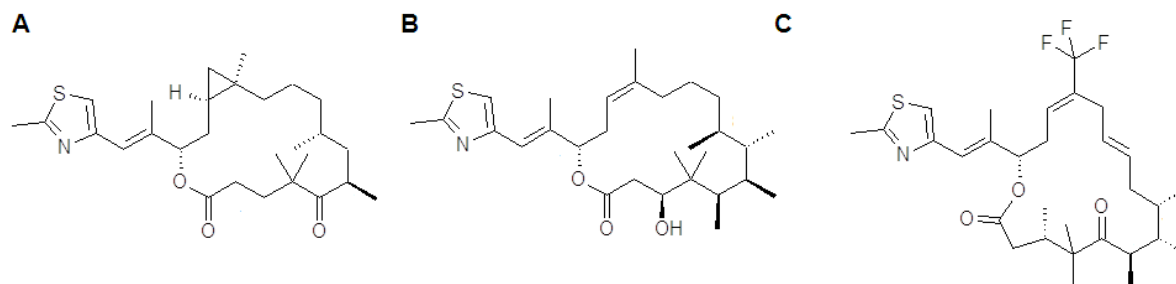


Figura 7.7. Estructuras de los *hits* computacionales resultantes del análisis. A. 69701800, B. 59072409, C. 59097730.

VIII. CONCLUSIONES

Los descriptores moleculares 2D permitieron generar modelos QSPR con herramientas quimioinformáticas, los cuales se adecuaron y validaron con métodos estadísticos. Los modelos lineales desarrollados son similares al modelo de Clark y mostraron ventajas teóricas en el análisis estadístico realizado.

El sistema heterotetramérico $\alpha_2\beta_2$ representa, de manera estable, el comportamiento microtubular en el intervalo de tiempo de simulación de MD. Aunque el sistema homodimérico es una simplificación, mostró ser suficiente para evaluar la estabilidad y las interacciones de los ligandos de unión a EPA con 80% de similitud estructural. Los modelos del sistema de tubulinas mostraron alta estabilidad cuando incluyen la interfase β - β , por lo que la interacción del asa M en la interfase resultó ser determinante.

En el tiempo de simulación de los sistemas homodiméricos, los ligandos estabilizadores microtubulares tuvieron la notable capacidad de incrementar el número de puentes de hidrógeno en la interfase β - β incluyendo el sistema control positivo de referencia con EPA, respecto al sistema sin ligandos acoplados, el control negativo. Los estimados de energías de unión variaron entre los dos sistemas y los dos métodos utilizados, debido fundamentalmente a los tiempos de aplicación y diferencias algorítmicas. Los ligandos comparten la mayoría de los subespacios farmacofóricos propuestos, y se analizaron sus interacciones en el sitio de unión en β -tubulina. Los ligandos con mejor comportamiento estabilizador son 69701800, 59072409 y 59097730.

IX. PERSPECTIVAS

Los *hits* computacionales propuestos son candidatos para la evaluación experimental en sistemas amortiguados de tubulinas. En cultivos celulares, se espera la actividad estabilizadora microtubular si la cinética de transporte predomina sobre la cinética de las enzimas de biotransformación, de manera similar a como se observa *in vivo* con la epotilona A. La actividad citostática o citotóxica dependerá de las constantes de afinidad de los ligandos por las tubulinas, así como la inactivación de estos para la reanudación de la dinámica microtubular. La evaluación en modelos animales de enfermedad de Alzheimer se llevará a cabo si los compuestos que se proponen en este trabajo como *hits* computacionales no perturban de manera importante la homeostasis celular.

X. ANEXO I

X.1. Cribado virtual

El cribado virtual (*virtual screening*) tiene como finalidad seleccionar, de las bases de datos, moléculas biológicamente activas frente a dianas particulares empleando herramientas computacionales. Si se conoce el ligando con actividad biológica, se puede realizar la búsqueda de estructuras con base en su similitud con el ligando conocido, a lo cual se le conoce como “Diseño basado en la estructura del ligando”.

X.1.1. Coeficiente de Tanimoto

En la búsqueda de estructuras basada en similitud, los sistemas quimioinformáticos utilizan una representación vectorial denominada *fingerprint* (huella dactilar), en donde una estructura molecular es representada por un vector y cuyos componentes son la presencia o ausencia de un determinado grupo funcional. Esta caracterización estructural es de gran ayuda para identificar, en bases de datos grandes, la presencia de alguna propiedad o una subestructura molecular.

Dadas dos estructuras moleculares A y B en un espacio topológico con vectores \vec{A} y \vec{B} respectivamente, el coeficiente de Tanimoto (Ecuación 1) para vectores binarios define la similitud entre estructuras.

$$Tan = 100 \frac{\sum_i A \cap B}{\sum_i A \Delta B} \quad (1)$$

El numerador denota la suma de la intersección; la suma de los elementos comunes a las especies A y B, y el denominador denota el tamaño de la diferencia simétrica; la suma de los elementos en la especie A y en la especie B pero no en ambas. *PubChem* genera un *fingerprint* subestructural binario de 881 bits para estructuras químicas destinado a la búsqueda de similitud molecular. Cada bit representa, por ejemplo, la presencia de un grupo funcional, el número de átomos de una subestructura, un tipo de sistema anular, átomos enlazantes, entorno atómico, etc., en una estructura química.²⁹

X.1.2. Similitud estructural

El coeficiente de Tanimoto porcentual define el porcentaje de similitud de dos estructuras, basado en el tipo de *fingerprint* utilizado. Los porcentajes de similitud con los que se puede realizar la búsqueda de estructuras similares en *PubChem* son mayores o iguales a 80%, 90% y 95%.²⁹

X.1.3. Similitud de interacción proteína-ligando (PLIFs)

Las interacciones proteína-ligando, en particular interacciones hidrofóbicas, puentes de hidrógeno convencionales y no convencionales, puentes salinos y residencia de moléculas de agua, en intervalos crecientes de 2 nm a 4 nm, también se representan con vectores binarios, y sus coeficientes de Tanimoto porcentuales definen la similitud de interacción proteína-ligando (*PLIF*). Las interacciones se consideran para un mismo residuo de aminoácido o para átomos individuales del residuo.

X.2. Relaciones cuantitativas de estructura-actividad (QSAR) y estructura-propiedad (QSPR)

Las relaciones cuantitativas de estructura actividad (*Quantitative structure-activity relationships, QSAR*) y estructura-propiedad (*Quantitative structure-property relationships, QSPR*) se emplean para diseñar compuestos más potentes o menos tóxicos a partir de una serie de candidatos. *QSAR/QSPR* se basa en que las propiedades físicas, químicas y biológicas son consecuencia de la distribución electrónica, y que pueden ser descritas en estos términos.⁷⁶ Cualquier cambio en la estructura molecular altera las propiedades fisicoquímicas y biológicas de las moléculas. Las propiedades cuantitativas y relacionadas con la distribución electrónica son los descriptores, y por simplicidad se busca que la relación entre los descriptores y la actividad biológica experimental sean lineales, mediante un análisis de regresión lineal.

X.2.1. Descriptores moleculares

La información estructural y propiedades fisicoquímicas se representan numéricamente en descriptores que codifican a las moléculas.⁷⁷ Los descriptores fisicoquímicos son ampliamente utilizados en estudios *QSAR*.

X.2.2. Ecuación de Clark

El paso de una molécula a través de la barrera hematoencefálica puede predecirse con la ecuación de Clark (Ecuación 2), la cual considera los descriptores *PSA* y *clogP*. *logBB* es una estimación del logaritmo del cociente de concentraciones encefálicas y plasmáticas.³²

$$\log BB = -0.0148 \text{ \AA}^{-2}(\pm 0.001)PSA + 0.152(\pm 0.036)clogP + 0.139(\pm 0.073) \quad (2)$$

El área superficial polar (*PSA*) se define como la suma de las superficies de átomos polares en una molécula, y presenta buena correlación con la información experimental de transporte biológico. Se ha utilizado exitosamente para la predicción de su absorción intestinal, penetración en monocapas de la línea celular Caco-2 y el cruce de la BBB. El área superficial polar topológica (*tPSA*) está basada

en la suma de contribuciones superficiales de fragmentos polares tabulados. El coeficiente de correlación de PSA con tPSA es 0.99 con $n = 34810$.⁷⁸

El logaritmo de la constante de reparto calculada (clogP) es una aproximación del valor experimental de la constante de reparto octanol-agua (Pow). El método de cálculo XLOGP3 (Ecuación 3) considera las contribuciones a,c y ocurrencias A,C de los átomos i,j en un análogo de un compuesto con valor de logP conocido ($\log P_0$)⁶¹ en un espacio métrico.

$$XLOGP3 = \log P_0 + \sum_{i=1}^M a_i(A_i - A_{i,0}) + \sum_{j=1}^N c_j(C_j - C_{j,0}) \quad (3)$$

X.2.3. Regresión lineal múltiple

El análisis de regresión es una técnica estadística para investigar y modelar la relación entre variables.⁷⁹ En estudios de QSAR se modela la relación de j descriptores con la actividad biológica experimental, en este caso el logBB.

El modelo de regresión lineal tiene la ecuación de una recta multivariada (Ecuación 4):

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + \varepsilon_R \quad (4)$$

Donde y es el regresando (variable dependiente) y x_j son los regresores (variables independientes). Los coeficientes β_j se estiman por el método de cuadrados mínimos. La adecuación del modelo se evalúa por los estadísticos coeficiente de correlación de Pearson r (Ecuación 5), coeficiente de determinación r^2 (Ecuación 6a) y coeficiente de determinación ajustado \bar{r}^2 (Ecuación 6b). El coeficiente de correlación es la covarianza normalizada del regresando y el regresor, e indica el grado de dependencia lineal. Geométricamente es el coseno del ángulo formado entre los dos vectores de N componentes. El coeficiente de determinación r^2 es una medida de la variabilidad de la variable dependiente logBB que queda después de haber tenido en consideración a x_j .⁸⁰ La variabilidad de logBB queda explicada en $100r^2\%$. El coeficiente de determinación ajustado \bar{r}^2 sólo aumenta al agregar una variable al modelo y es muy utilizado en la regresión lineal múltiple.

$$r = \frac{SC_{xy}}{\sqrt{SC_x SC_y}} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (5)$$

$$r^2 = 1 - \frac{SC_{res}}{SC_T} \quad (6a)$$

$$\bar{r}^2 = 1 - \left[\frac{(1 - r^2)(n - 1)}{n - k} \right] \quad (6b)$$

Los residuales (Ecuación 7) son la diferencia entre el valor observado y el valor ajustado o predicho.

$$e_i = y_i - \hat{y}_i \quad (7)$$

X.2.3.1. Colinealidad

Si existe una correlación lineal entre los descriptores empleados en el modelo de regresión lineal múltiple, el ajuste cambia de manera errática ante cambios mínimos de los datos. Una forma de evaluar la colinealidad de los regresores es emplear la matriz de correlación R (Ecuación 8), donde cada elemento r_{ij} es el coeficiente de correlación entre las dos variables independientes. Si $r_{ij} > 0.6$, se debe excluir uno de los descriptores antes de realizar la regresión.

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} & r_{1j} \\ r_{21} & 1 & r_{23} & r_{2j} \\ r_{31} & r_{32} & 1 & r_{3j} \\ r_{i1} & r_{i2} & r_{i3} & 1 \end{pmatrix} \quad (8)$$

X.2.3.2. Prueba de significancia de la regresión

Esta prueba se utiliza para determinar si hay una relación lineal entre la respuesta y cualquiera de las variables regresoras x_j .⁸¹ La prueba puede ser global o sobre coeficientes individuales.

X.2.3.3. Grados de libertad

Si la muestra se representa como un punto en el espacio métrico de N dimensiones con media poblacional en el origen, los grados de libertad están relacionados con la movilidad permitida de ese punto, que corresponde con la desviación de la media muestral respecto a la media poblacional. Para localizar el punto muestral en una trayectoria, se define una relación que secciona el hiperespacio para restringir su movimiento a un hiperespacio de menor dimensión. Los grados de libertad representan la dimensión hiperespacial necesaria para que el punto muestral quede fijo a la condición impuesta por la relación. En la Figura 10.1 se ejemplifica un espacio métrico con $N=3$ y la interpretación geométrica de los grados de libertad. Los puntos muestrales $A(a_1, a_2, a_3)$ y $B(b_1, b_2, b_3)$ tienen la misma media muestral, donde forman un subespacio de dimensión $N-1$. Los círculos en el plano color azul forman el espacio muestral. El origen $O(0,0,0)$ es la media poblacional, el cual genera la recta OM que es ortogonal al plano en $m(m_1, m_2, m_3)$.

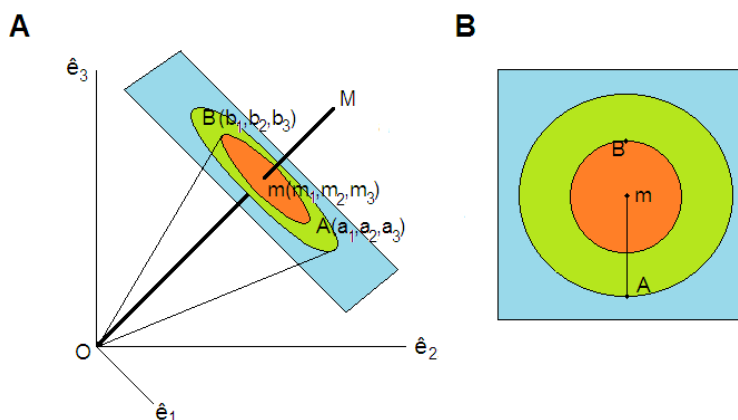


Figura 10.1. Representación geométrica de los grados de libertad. El espacio poblacional tiene dimensión $N=3$ y el espacio muestral tiene dimensión $N-1=2$, definido por el plano que forman las medias muestrales. Los grados de libertad son la dimensión resultante de acotar el espacio poblacional con la relación impuesta por las medias muestrales.

En la Figura 10.1, los segmentos de recta son:

$$\begin{aligned}\overline{OA} &= (a_1 - 0) + (a_2 - 0) + (a_3 - 0) = \sum (a - \mu), \overline{OA^2} = \sum (a - \mu)^2 \\ \overline{Om} &= (m_1 - 0) + (m_2 - 0) + (m_3 - 0) = N(m - \mu), \overline{Om^2} = N^2(m - \mu)^2 \\ \overline{Am^2} &= \sum (a - \mu)^2 - N^2(m - \mu)^2 = \sum a^2 - N^2m^2 = N^2s^2\end{aligned}$$

El cociente $\frac{Om}{Am} = \frac{m - \mu}{s}$, la cotangente del ángulo formado entre el vector de la media poblacional y el vector de la muestra, es proporcional a cuánto se desvía una media muestral respecto a la media poblacional debido a su error estándar. La fluctuación de esta razón de muestra a muestra produce la distribución t de Student. Las muestras sucesivas se ubican en su mayoría alrededor del punto m , en el círculo anaranjado. La minoría de muestras se ubica en posiciones radiales alejándose de m , en el círculo verde.

X.2.3.4. Prueba global de significancia

Una hipótesis estadística es una afirmación acerca de la distribución de una o más variables aleatorias.⁸² Una hipótesis debe especificar no sólo la forma funcional de la distribución subyacente, sino también los valores de todos los parámetros. Si las suposiciones establecen que no hay diferencia en los parámetros, se trata de una hipótesis nula H_0 . Cualquier otra suposición constituye una hipótesis alterna H_1 . Una hipótesis estadística permite calcular un estadístico de prueba, que está dividido en dos subconjuntos, una región de aceptación de H_0 y una región de rechazo de H_0 . El rechazo de H_0 cuando es verdadera se llama error tipo I, y la probabilidad de cometerlo es α , o nivel de significancia.⁸²

La prueba de significancia contrasta a H_0 y H_1 y es una generalización del análisis de varianza (Ecuación 9). De manera general (Tabla 10.1), las hipótesis nulas afirman que la media poblacional de donde se tomó la muestra (μ_A) o que la diferencia de dos medias poblacionales ($\mu_A - \mu_B$) tiene relación con un valor a . Las hipótesis alternas consideran las demás posibilidades de valor de la media o relación entre las medias.

Tabla 10.1. Hipótesis estadísticas concernientes a las medias

$H_0: \mu_A = a, \mu_A - \mu_B = 0$	Hipótesis a probar
$H_1: \mu_A \neq a, \mu_A - \mu_B \neq 0$	Prueba bilateral o a dos colas
$H_1: \mu_A < a, \mu_A - \mu_B < 0$	Prueba unilateral o a una cola
$H_1: \mu_A > a, \mu_A - \mu_B > 0$	Prueba unilateral o a una cola

Para el análisis de regresión lineal:

$$H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_j = 0$$

$$H_1: \beta_j \neq 0 \text{ al menos para una } k$$

$$SC_T = SC_R + SC_{res} \quad (9)$$

El estadístico F_c (Ecuación 10) tiene una distribución $F_{\alpha, k-1, n-k} = F_t$ (Ecuación 11).

$$F_c = \frac{\frac{SC_R}{k-1}}{\frac{SC_{res}}{n-k}} = \frac{CM_R}{CM_{res}} \quad (10)$$

$$\int_0^x \frac{\Gamma\left(\frac{v_R + v_{res}}{2}\right) \left(\frac{v_R}{v_{res}}\right)^{\frac{v_R}{2}} x^{\frac{v_R}{2}-1}}{\Gamma\left(\frac{v_R}{2}\right) \Gamma\left(\frac{v_{res}}{2}\right) \left(1 + \frac{v_R x}{v_{res}}\right)^{\frac{v_R + v_{res}}{2}}} dx = 1 - \alpha \quad (11)$$

Donde el integrando es la función de densidad de probabilidad de Fisher-Snedecor (Figura 10.2) y F_t es el valor de x que satisface la ecuación 11 con $v_R = k-1$ y $v_{res} = n-k$ grados de libertad. En este caso, $1 - \alpha$ representa el nivel de confianza de la prueba, usualmente de 0.95 (95%), y $\Gamma(z)$ es la función gamma (Ecuación 12).

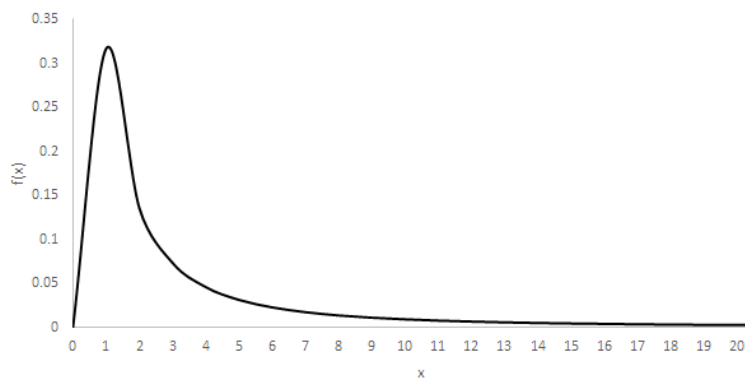


Figura 10.2. Densidad de probabilidad F de Fisher-Snedecor con $v_R=6$ y $v_{res}=2$ grados de libertad.

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx \quad (12)$$

La prueba compara el estadístico de prueba F_c con el valor de la variable F_t con esta densidad de probabilidad. Si $F_c > F_t$, se rechaza H_0 . El procedimiento se resume en una tabla de análisis de varianza (Tabla 10.2).

Tabla 10.2. Tabla del análisis de varianza

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_c
Regresión	SC_R	$k-1$	CM_R	$\frac{CM_R}{CM_{res}}$
Residuales	SC_{res}	$n-k$	CM_{res}	
Total	SC_T	$n-1$		

El valor p (Ecuación 13) es la probabilidad, bajo la función de densidad $f(x)$, de que los estadísticos de una prueba puedan ser iguales o más extremos que el resultado obtenido.⁸³ Cuando $\alpha=0.05$, la prueba es significativa cuando $p < 0.05$ ($H_0: \mu_A - \mu_B > 0$ o $\mu_A - \mu_B < 0$). Si la prueba es a bilateral ($H_0: \mu_A - \mu_B \neq 0$), el resultado de (13) se deben sumar las integrales definidas en los intervalos $(-\infty, -x_c]$ y $[+x_c, \infty)$, donde x_c es el valor del estadístico de prueba.

$$p = \int_{+x_c}^{\infty} f(x) dx \quad (13)$$

X.2.3.5. Prueba de significancia a coeficientes individuales

Las hipótesis para probar la significancia de cualquier coeficiente individual de regresión β_j son:⁶²

$$H_0: \beta_j=0$$

$$H_1: \beta_j \neq 0$$

Si no se rechaza H_0 se puede eliminar el regresor x_j del modelo. El estadístico de prueba es t_c (Ecuación 14):

$$t_c = \frac{\beta_j}{EE(\beta_j)} \quad (14)$$

Donde $EE(\beta_j)$ es el error estándar del regresor estimado.

El estadístico t_c se contrasta con el valor t de tablas t_t (Ecuación 15):

$$\int_{-\infty}^t \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt = 1 - \alpha \quad (15)$$

Donde el integrando es la función de densidad de probabilidad t de Student (Figura 10.3) y t_t es el valor de t que satisface la ecuación 15 con $\nu=n-k$ grados de libertad.

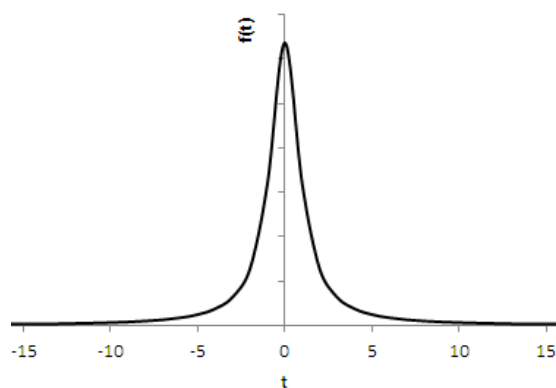


Figura 10.3. Densidad de probabilidad t de Student con $\nu=1$ grado de libertad.

X.2.4. Adecuación del modelo

Un residual es la desviación entre los datos y el ajuste, y una medida de la variabilidad de respuesta que no explica el modelo de regresión.⁶³ Los residuales escalados son útiles para determinar observaciones atípicas (*outliers*). Algunos métodos para escalar residuales se describen a continuación.

X.2.4.1. Residuales estandarizados (d_i)

Los residuales estandarizados (Ecuación 16) tienen esperanza cero y varianza aproximadamente unitaria. En consecuencia, un residual estandarizado grande ($d_i > 2$) indica que se trata de un valor atípico potencial.

$$d_i = \frac{e_i}{\sqrt{CM_{res}}} \quad (16)$$

X.2.4.2. Residuales estudentizados (r_i)

Si se usa CM_{res} como la varianza del i -ésimo residual e_i , sólo se tendrá una aproximación. Si se divide e_i por la desviación estándar exacta del i -ésimo residual se puede mejorar el escalamiento.⁶³ La ecuación 4 se puede escribir en notación matricial como:

$$Y = XB + E_R \quad (17)$$

Donde:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1j} \\ 1 & x_{21} & x_{22} & \cdots & x_{2j} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nj} \end{pmatrix}, \quad B = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_j \end{pmatrix}, \quad E_R = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Sean las matrices $C = (X^T X)^{-1}$ de $j \times j$ y $H = X(X^T X)^{-1} X^T$ la matriz de sombrero de $i \times i$. El elemento c_{jj} es el elemento diagonal jj de C y h_{ii} es el elemento diagonal ii de H . Los residuales estudentizados (Ecuación 18) se definen como:

$$r_i = \frac{e_i}{\sqrt{CM_{res}(1 - h_{ii})}} \quad (18)$$

X.2.4.3. Distribución de los residuales

Se espera que los residuales tengan una distribución aproximadamente normal. Para verificar esta asunción existen pruebas de normalidad, como la prueba de Shapiro-Wilk o estadístico W .

(a) Prueba de Shapiro-Wilk. Sean e_1, \dots, e_i, e_n los residuales independientes y de distribución aleatoria, y los valores $e_{(1)}, \dots, e_{(i)}, e_{(n)}$ son sus magnitudes ordenadas. El estadístico W (Ecuación 19)⁸⁴ de

Shapiro-Wilk evalúa la distribución de los residuales, y la asunción de normalidad se rechaza cuando $p < \alpha$ (comúnmente $\alpha = 0.05$).

$$W = \frac{(\sum_{i=1}^n a_i e_{(i)})^2}{\sum_{i=1}^n (e_i - \bar{e})^2}, \quad W \in [0,1] \quad (19)$$

Donde a_i son constantes provenientes de las matrices de valor esperado y covarianza de la distribución normal.⁸⁵

X.2.4.4. Intervalos de confianza para los coeficientes de regresión

Un intervalo de confianza de $100(1-\alpha)\%$ (Ecuación 20) para el coeficiente β_j se define como:

$$\hat{\beta}_j - t_{\frac{\alpha}{2}, \nu} CM_{res} \sqrt{c_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\frac{\alpha}{2}, \nu} CM_{res} \sqrt{c_{jj}} \quad (20)$$

Donde:

$$EE(\beta_j) = CM_{res} \sqrt{c_{jj}}$$

X.2.4.5. Distancia de Cook

El escalamiento del residual generado por la eliminación de la i -ésima observación eliminada se conoce como escalamiento externo, y emplea un estimado de σ^2 generado internamente. El estimado de σ^2 generado externamente se denota por $s_{i(-i)}^2$ (Ecuación 21).

$$s_{i(-i)}^2 = \frac{(n - k - 1)CM_{res} - \frac{e_i^2}{1 - h_{ii}}}{n - k - 1} \quad (21)$$

La distancia de Cook D_i (Ecuación 22) estima la influencia para el modelo del i -ésimo valor predicho.

$$D_i = \frac{h_{ii} e_i^2}{s_{i(-i)}^2 (k - 1) (1 - h_{ii})^2} \quad (22)$$

Las distancias de Cook siguen una distribución beta con parámetros a y b que dependen de las n observaciones y los j descriptores. El cálculo de los parámetros a y b de la distribución de D_i implica transformaciones jacobianas a partir de la distribución F ,⁶⁴ por lo que usualmente se emplea esta última para la interpretación de los gráficos de D_i , siendo valores atípicos las observaciones con $D_i > F_{0.5, n, n-k}$. Otro criterio extendido es el uso del valor $4/n-k$ ⁸⁶ como mínimo de D_i para considerar a las observaciones como valores atípicos.

X.2.5. Validación de la regresión

X.2.5.1. Validación interna

Los datos de la variable dependiente logBB pueden dividirse razonablemente en dos grupos, los datos de estimación (entrenamiento) y los datos de predicción (prueba). Los datos de estimación se usan para formar el modelo de regresión y a continuación los datos de predicción se usan para estudiar la capacidad predictiva del modelo. A esta división de los datos se le llama validación cruzada interna. El estadístico PRESS (*prediction error sum of squares*) es una forma de división de datos, y el estadístico q^2 mide en forma aproximada cuánto cabe esperar que el modelo explique la variabilidad de las nuevas observaciones.

PRESS (Ecuación 23) es la suma de cuadrados de la diferencia entre el valor observado y el valor predicho de la variable dependiente logBB, donde $\hat{y}_{i(-i)}$ es el estimado \hat{y}_i del modelo cuando el valor i -ésimo ha sido removido de la regresión (*leave-one-out, LOO*).

El parámetro q^2 es equivalente a r^2 , y emplea el estadístico PRESS en lugar de la suma de cuadrados de residuales (Ecuación 24).

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_{i(-i)})^2 \quad (23)$$

$$q_{int}^2 = 1 - \frac{PRESS}{SC_T} \quad (24)$$

Los coeficientes de determinación en la validación externa son el coeficiente de determinación entre el valor predicho y observado (r_{val}^2 , ecuación 25), y los coeficientes de determinación ajustados al origen para el valor predicho contra el valor observado (r_0^2 , ecuación 26a) y para el valor observado contra el valor predicho (r'_0^2 , ecuación 26b). Los valores de los coeficientes de determinación en la validación deben ser mayores a 0.6 para un buen modelo.^{65,67}

$$r_{val}^2 = 1 - \frac{\sum_{i=1}^n (y_{i,obs} - \hat{y}_i)^2}{\sum_{i=1}^n (y_{i,obs} - \bar{y}_{i,obs})^2} \quad (25)$$

$$r_0^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_{i,0})^2}{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2} \quad (26a)$$

$$r'_0^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{i,0})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (26b)$$

XI. ANEXO II

XI.1. Modelado de proteínas

Debido a la naturaleza desordenada de varios segmentos de las tubulinas, las estructuras no pueden cristalizarse de manera íntegra, por lo que se recurre a los métodos de modelado. El modelado puede ser *de novo*, a partir de la estructura primaria del polipéptido, o por homología, cuando se encuentran disponibles estructuras cristalizadas que funcionan como template y la secuencia de interés. Para que el modelo sea aceptable debe existir una alta similitud entre las estructuras primarias, lo cual se puede realizar con herramientas computacionales. Los modelados estructurales deben validarse finamente con métodos estadísticos y empíricos, como el valor z , QMEAN (análisis del modelo cualitativo de energía) y análisis conformacional. El valor z (ecuación 27) es el estadístico z_c que infiere la confiabilidad del modelo a partir de una población normal de estructuras resueltas (Figura 11.1, ecuación 28). QMEAN considera las posiciones entre carbonos β y todos los átomos, potencial de solvatación y potencial de torsión. El potencial de torsión analiza los ángulos ϕ (formados entre los planos generados por los átomos C-N-C α y N-C α -C) y ψ (formados entre los planos generados por los átomos N-C α -C y C α -C-N) cada tres residuos de aminoácidos adyacentes (Figura 11.2). El diagrama de Ramachandran establece la naturaleza de las conformaciones dependiendo de los valores de ϕ y ψ , prediciendo las estructuras secundarias de hojas beta, hélices dextrógiras y hélices levógiras.

$$z_c = \frac{\bar{d} - \mu_d}{\sigma_d} \quad (27)$$

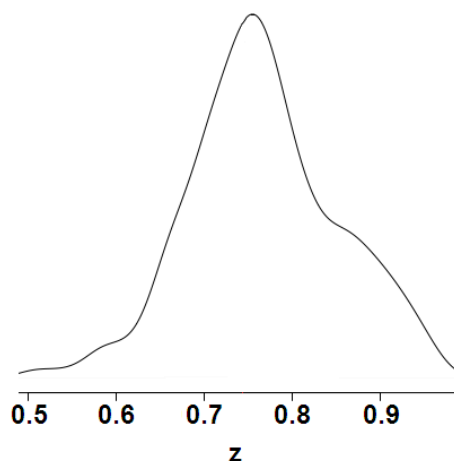


Figura 11.1. Densidad aproximadamente normal generada a partir del grupo de estructuras proteicas resueltas, con la cual se contrasta el valor z_c (Modificado de SWISS-MODEL).⁴⁰

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (28)$$

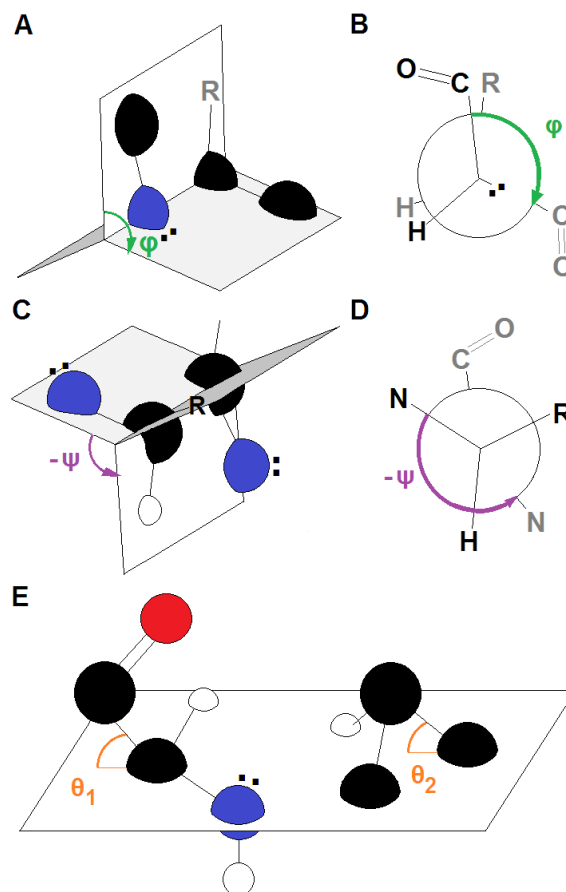


Figura 11.2. Ángulos diedros. (A) El ángulo φ es el ángulo formado entre los planos generados por los átomos C-N-C α y N-C α -C. (B) Proyección de Newman de una conformación anticlinal con $\varphi > 0$. (C) El ángulo ψ es el ángulo formado entre los planos generados por los átomos N-C α -C y C α -C-N. (D) Proyección de Newman de una conformación antiperiplanar con $\psi < 0$. (E) Ejemplos de ángulos diedros impropios, entre el plano formado por tres átomos enlazados a un cuarto átomo.

XI.1.2. Análisis cualitativo del modelo de energía (QMEAN)

QMEAN es una función de evaluación con tres potenciales estadísticos y dos términos adicionales correspondientes con la concordancia de estructuras secundarias modeladas y de referencia (*SSE, second-structure experimental*) y accesibilidad al solvente (*ACC*)⁴², (Tabla 11.1) La referencia es un conjunto de estructuras resueltas con resolución alta depositadas en el PDB. Los potenciales estadísticos son el potencial pareado dependiente de la distancia, el potencial pareado de nivel de residuo y el potencial basado en C β .

El potencial pareado dependiente de la distancia considera pares en contacto de al menos cuatro residuos de aminoácidos. Los contactos $C\alpha$ y $C\beta$ se consideran los sitios de interacción.

El potencial pareado de nivel de residuo se calcula con base en la frecuencia de contactos extraída de residuos con la misma estructura secundaria.

El potencial basado en $C\beta$ utiliza el potencial referente a estructura secundaria con datos empíricos.

El potencial de solvatación cuenta los centros de interacción en $C\beta$ dentro de una esfera de $r=9 \text{ \AA}$ para cada residuo. La accesibilidad relativa se calcula como el cociente del conteo en la estructura analizada y el conteo máximo en las estructuras de referencia. El potencial de solvatación refleja la propensión de un residuo para una accesibilidad de solvente comparada con cualquier otro residuo.

El potencial de ángulo de torsión refleja la propensión de cierto residuo para una torsión dada con cualquier otro residuo. La descripción de la geometría local para un residuo incluye la de residuos adyacentes, incluye los ángulos diedros φ_{i-1} , ψ_{i-1} , φ_i , ψ_i , φ_{i+1} , ψ_{i+1} para el i -ésimo residuo.

A los potenciales calculados se les calcula el valor z para compararlos con los valores z de la estructura de referencia.

Tabla 11.1. Evaluaciones realizadas por QMEAN

Potenciales pareados		Potenciales de solvatación	Potenciales de torsión	Potenciales de accesibilidad al solvente	Potenciales de estructura secundaria predicha
Interacciones $C\beta$	Interacciones de todos los átomos				

XI.2. Parametrización

La parametrización de moléculas pequeñas, los ligandos, consiste en la creación de métricas llamadas topologías asignando los tipos de átomos, cargas parciales, distancias de enlaces, ángulos diedros propios e impropios; y la creación de parámetros y constantes del campo de fuerza, como constantes de interacciones.⁴⁵ Las cargas parciales se calculan de acuerdo al procedimiento establecido en cada MMFF.⁸⁷

XI.3. Acoplamiento molecular

El acoplamiento molecular (*docking*) es una aproximación *in silico* a los modos de unión y la estabilidad de complejos interactuantes, frecuentemente ligandos pequeños con macromoléculas y proteína-proteína. El acoplamiento puede ser ciego, cuando se permite que el ligando pueda unirse a cualquier región de la macromolécula, o dirigido, cuando se define una superficie específica o cavidad frecuentemente basada en información experimental o cristalográfica. La rejilla de búsqueda (*grid*) se genera en el sitio de unión o posicionando un átomo sonda en cada vértice. Los estudios de acoplamiento calculan una SF (*scoring function*) que representa una aproximación al ΔG de unión del ligando.⁸⁸ La función SF puede estar basada en un campo de fuerza (*force field*, FF), ser empírica o basada en el conocimiento.

XI.3.1. SF basadas en campos de fuerza

A partir de mecánica molecular, se aproxima el ΔG_{bind} a partir de la suma de interacciones electrostáticas, van der Waals (a través de la aproximación del potencial de Lennard-Jones) y de enlaces de hidrógeno⁷⁷ (ecuación 29).

$$SF = \frac{1}{4\pi\epsilon_0} \sum_{ij}^n \frac{q_i q_j}{\|\vec{r}\|} + \sum_{ij}^n \epsilon \left[\left(\frac{r_{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{min}}{r_{ij}} \right)^6 \right] + \sum_k^n E_{hbond,k} \quad (29)$$

Donde $E_{p,L-J} = \sum_{ij}^n \epsilon \left[\left(\frac{r_{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{min}}{r_{ij}} \right)^6 \right]$ es la energía potencial de Lennard-Jones.

XI.3.2. SF empíricas o de enfoque bayesiano

Estiman ΔG_{bind} sumando términos de interacción derivados de la contribución ponderada de parámetros estructurales como número de enlaces de hidrógeno, interacciones iónicas, interacciones hidrofóbicas, entropía, etc. La ponderación de cada parámetro se obtiene por ajuste a constantes de unión experimentales de un conjunto de complejos proteína-ligando.⁷⁷

XI.3.3. SF basadas en el conocimiento o de enfoque frecuentista

Representan la afinidad como suma de interacciones de pares de átomos proteína-ligando. Estos potenciales se derivan a partir de complejos de estructuras conocidas del PDB, donde las distribuciones de probabilidad de las distancias interatómicas entre diferentes pares de tipos de átomos proteína-ligando se convierten en funciones de potencial, asumiendo distribuciones energéticas dadas por la distribución de Boltzmann (ecuación 30). La energía libre de interacción se calcula sumando las contribuciones de los pares de átomos de una cierta distancia.⁷⁷

$$p_i = \frac{e^{-\frac{E_i}{k_B T}}}{\sum_{j=1}^N e^{-\frac{E_j}{k_B T}}} \quad (30)$$

Donde el denominador es la función de partición (Q), y la energía del sistema está dada por la esperanza matemática de la función de densidad (ecuación 31).

$$\langle E \rangle = E(E) = \frac{\sum_{j=1}^N E_j e^{-\frac{E_j}{k_B T}}}{\sum_{j=1}^N e^{-\frac{E_j}{k_B T}}} \quad (31)$$

XI.3.4. SF en programas de acoplamiento molecular

XI.3.4.1. SF en AutoDock Vina

La función de evaluación en AutoDock Vina es una combinación de SF basadas en el conocimiento y empíricas, y su cálculo emplea menor tiempo computacional que la SF de AutoDock 4.2.⁴⁹ La SF implementada en Vina consiste en el cálculo de cinco contribuciones (Ecuación 32), tres de ellas estéricas (Ecuaciones 33 a 35), una hidrofóbica (Ecuación 36) y de enlace de hidrógeno (Ecuación 37) para los tipos de átomos i y j .

$$SF = \sum_{i>j} f_{titj}(d_{ij} = r_{ij} - r_{vdw,ti} - r_{vdw,tj}) \quad (32)$$

$$gauss_1(d_{ij}) = e^{-\left(\frac{d_{ij}}{0.5 \text{ \AA}}\right)^2} \quad (33)$$

$$gauss_2(d_{ij}) = e^{-\left(\frac{d_{ij}-3 \text{ \AA}}{2 \text{ \AA}}\right)^2} \quad (34)$$

$$rep(d_{ij}) = \begin{cases} d_{ij}^2, & \text{si } d_{ij} < 0 \\ 0, & \text{si } d_{ij} \geq 0 \end{cases} \quad (35)$$

$$hp(d_{ij}) = \begin{cases} 1, & \text{si } d_{ij} < 0.5 \text{ \AA} \\ 0, & \text{si } d_{ij} > 1.5 \text{ \AA} \end{cases} \quad (36)$$

$$hbond(d_{ij}) = \begin{cases} 1, & \text{si } d_{ij} < -0.7 \text{ \AA} \\ 0, & \text{si } d_{ij} > 0 \end{cases} \quad (37)$$

Para las contribuciones hidrofóbica (hp) y de enlace de hidrógeno ($hbond$), los valores dentro del intervalo que define cada función se obtienen por interpolación lineal.

XI.3.5. Algoritmos genéticos

Un método para generar los confórmeros de los ligandos son los algoritmos genéticos (GA). El arreglo particular de un ligando y una proteína puede ser definido por un grupo de valores que describen la traslación, la orientación y la conformación de ligando respecto a la proteína en dos vectores de n dimensiones. Estos valores son las variables de estado, y cada variable de estado corresponde a un gen. El estado del ligando corresponde al genotipo, y sus coordenadas atómicas en el sitio son el fenotipo. Parte de la descendencia experimenta una mutación, en la cual los genes cambian en una cantidad aleatoria.^{89,90} En el algoritmo genético lamarckiano las mutaciones se pueden realizar añadiendo un número real aleatorio que tiene una distribución de Cauchy (ecuación 38).

$$C(x, x_0, \gamma) = \frac{\gamma}{\pi[\gamma^2 + (x - x_0)^2]} \quad (38)$$

Donde x_0 es el corrimiento que especifica la moda de la distribución y γ es el parámetro de escala que afecta el ancho medio. Generalmente se usa $x_0 = 0$. El número aleatorio generado con esta distribución se adiciona a los genes correspondientes para el vector unitario de orientación y de ángulos diedros para traslación y conformación. El ancho medio de la distribución de Cauchy ($x_0 = 0$, $\gamma = 1$) es mayor que en la distribución normal ($\mu = 0$, $\sigma = 1$), por lo que los números aleatorios generados permiten una mutación más grande y limita la convergencia prematura en un mínimo local (Figura 11.3), y una búsqueda más rápida de éste.⁹¹

Para seleccionar los confórmeros a los que se les generarán mutaciones sucesivas, se utiliza una relación que evalúa su ajuste (ecuación 39).⁸⁹

$$n = \frac{f_{MÁX}(E) - f_i(E)}{f_{MÁX}(E) - \langle f(E) \rangle} \quad (39)$$

Donde n es el confórmero a seleccionar, $f_{MÁX}(E)$ es el ajuste de mayor valor energético, $f_i(E)$ es el ajuste del individuo y $\langle f(E) \rangle$ es el ajuste promedio.

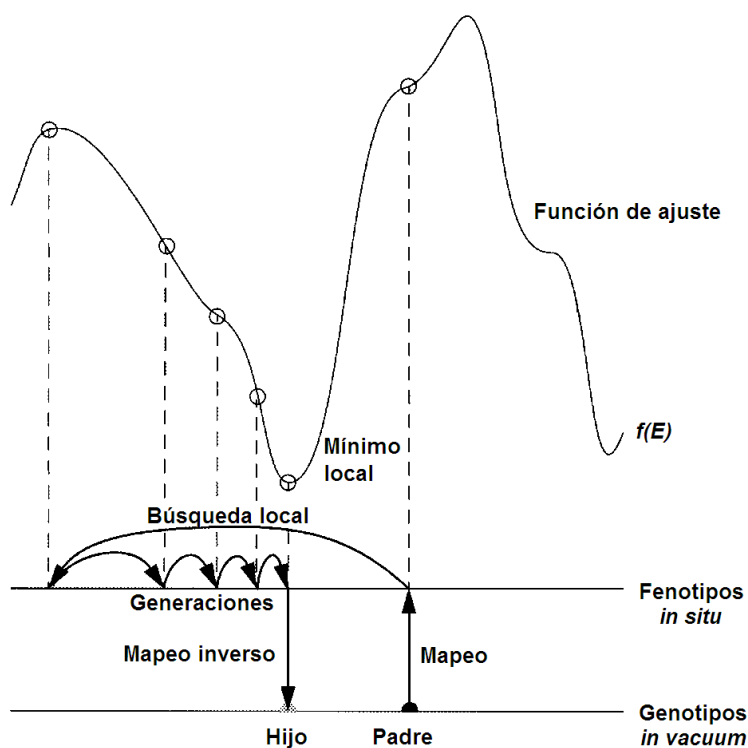


Figura 11.3. Generación de mutaciones mediante el algoritmo genético lamarckiano. El confórmero padre (genotipo) se genera sin considerar el sitio de unión, y posteriormente se acopla (fenotipo). Se generan las mutaciones a la conformación mediante la adición del número aleatorio con distribución de Cauchy a los vectores orientacional y angular diédrico, convergiendo a un mínimo local de la función de ajuste (Modificado de Morris, et al).⁸⁹

XI.4. Dinámica molecular

La simulación de dinámica molecular (MD) aproxima la estabilidad y el comportamiento de moléculas a través de un intervalo de tiempo determinado. La simulación MD resuelve las ecuaciones clásicas de movimiento (ecuación 40) de N átomos interactuantes por un método numérico.

$$-\nabla E_{p,i} = F_i = m_i \frac{\partial^2 r_i}{\partial t^2} \quad (40)$$

XI.4.1. Campo de fuerza (FF)

El campo de fuerza define el tratamiento de átomos y enlaces. Asigna a cada átomo un tipo (*atom type*) en función de su hibridación, carga y átomos enlazantes. A cada *atom type* le corresponde un grupo de parámetros (constantes de fuerza, radios atómicos, carga, masa, etc.). Cada FF define una ecuación de energía potencial.⁷⁷

XI.4.1.1. CHARMM

El campo de fuerza CHARMM está enfocado en biomoléculas y usa funciones de energía clásicas (empíricas y semiempíricas) y cuánticas.⁴⁶ La función de energía potencial es la suma de contribuciones internas y no enlazantes, que incluye los términos de enlace, ángulo de valencia, de Urey-Bradley, ángulos diedros, ángulos impropios y correcciones torsionales.

XI.4.1.2. GROMOS

El campo de fuerza GROMOS se aplica a biomoléculas y emplea métodos de mecánica molecular, dinámica estocástica y minimización de energía.⁹²

XI.4.2. Minimización de la energía

En la búsqueda de soluciones a la ecuación de energía potencial en función de las coordenadas atómicas, se encuentran los mínimos energéticos, donde $-\nabla E_{p,i}$ es cero. Los métodos más utilizados en mecánica molecular son los métodos iterativos *Steepest Descent* y gradiente conjugado.⁷⁷ Deben especificarse las condiciones de la minimización de la energía (Tabla 11.2).

Tabla 11.2. Condiciones de la minimización de la energía del sistema

Variable	Valor o argumento
Radio iónico de corto alcance	1.0 nm
Radio de van der Waals de corto alcance	1.0 nm
Magnitud de la fuerza máxima	<1000 kJnm/mol
Solvente	Explícito, agua con Na ⁺ y Cl ⁻ , 0.15 M

XI.4.2.1. *Steepest Descent* (SD)

La dirección de descenso (ecuación 41) corresponde a la posición donde el potencial sea menor. El máximo desplazamiento d_{MAX} es definido inicialmente.

$$r_{n+1} = r_n + d_{MAX} \frac{F_n(r)}{|F_n(r)|_{MAX}} \quad (41)$$

Si $V_{n+1} < V_n$, se acepta la nueva posición.

Si $V_{n+1} \geq V_n$ se rechaza la nueva posición.

XI.4.2.2. Gradiente conjugado (CG)

La dirección de búsqueda (ecuaciones 42a y 42b) se establece a partir del gradiente actual y del gradiente del paso anterior. El conjunto de direcciones generado no es ortogonal y se evita el comportamiento oscilatorio en las cercanías del mínimo que caracteriza a *SD*, y convergiendo más rápido que éste. *CG* no se debe usar con restricciones en el sistema.⁹³

$$r_{n+1} = r_n + \nabla_n \quad (42a)$$

$$\nabla_n = -\nabla f(r_n) + g\nabla_{n-1} \quad (42b)$$

Donde *g* es una relación entre gradientes.

XI.4.3. Ensamble o colectivo canónico (NVT)

Para la simulación de NVT, las condiciones especificadas (Tabla 11.3) determinan la asignación de las velocidades iniciales y la temperatura del sistema.

Tabla 11.3. Condiciones de NVT

Variable	Valor o argumento
Integrador	<i>Leap-frog</i>
Número de pasos	5000
Paso de tiempo	2 fs = 0.002 ps
Tiempo de simulación	10 ps
Asignación de velocidades	Distribución de Maxwell-Boltzmann
Restricciones	Holonómicas
Temperatura de referencia (T_0)	300 K

XI.4.3.1. Método de integración de *Leap-frog*

Es un método iterativo para el problema con valores iniciales $r(0)$ de la ecuación 40. Con $r(0)$ y Δt , que depende del paso de tiempo, se calcula una velocidad inicial $v(0)$ (Ecuación 43a) a partir de ecuaciones iterativas para posición, velocidad y aceleración (Ecuaciones 43b a 43d).⁹⁴

$$\vec{v}_0 = \frac{\vec{a}_0 \Delta t}{2} \quad (43a)$$

$$\vec{r}_{k+1} = \vec{r}_k + \vec{v}_k \Delta t \quad (43b)$$

$$\vec{a}_{k+1} = -\nabla E_{p,k+1} \quad (43c)$$

$$\vec{v}_{k+1} = \vec{v}_k + \vec{a}_{k+1} \Delta t \quad (43d)$$

XI.4.3.2. Distribución de Maxwell-Boltzmann

En el modelo del gas ideal, las velocidades de las partículas puntuales son descritas por la densidad de probabilidad de velocidades de Maxwell-Boltzmann⁹⁵ (Ecuación 44).

$$f(\|\vec{v}\|) = \sqrt{\frac{2}{\pi}} \frac{\|\vec{v}\|^2 e^{-\frac{m\|\vec{v}\|^2}{2k_B T}}}{\left(\frac{k_B T}{m}\right)^{\frac{3}{2}}} \quad (44)$$

Donde la velocidad media es la esperanza de v (Ecuación 45).

$$\langle \|\vec{v}\| \rangle = E(\|\vec{v}\|) = \frac{\int_0^{\max\{\|\vec{v}\|\}} \|\vec{v}\| f(\|\vec{v}\|) d\|\vec{v}\|}{\int_0^{\max\{\|\vec{v}\|\}} f(\|\vec{v}\|) d\|\vec{v}\|} \quad (45)$$

XI.4.3.3. Restricciones holonómicas (L)

Cuando la distancia entre las coordenadas de dos átomos enlazados es invariante con la traslación y rotación de estos, se dice que se encuentran en un espacio homogéneo e isotrópico, el espacio euclídeo. En un espacio euclídeo con una dimensión adicional de tiempo, donde N partículas se describen con sus componentes de posición y velocidad, las restricciones holonómicas reducen L veces los grados de libertad del sistema, siendo los grados de libertad del sistema restringido $6N - L$. Cada restricción holonómica define un subespacio donde los posibles movimientos del sistema pueden ocurrir. L aplica una fuerza de restricción en cada partícula, para mantenerla en el subespacio correspondiente.⁹⁶ Se restringen sólo las posiciones de las cadenas principales de los polipéptidos, simulando una fuerza que limita su movimiento.⁹³

XI.4.3.4. Cálculo de la temperatura

La temperatura del sistema se calcula con la media colectiva de la energía cinética (Ecuación 46).

$$T = \frac{2}{3N_d f k_B} \left\langle \sum_{i=1}^N m_i \|v_i\|^2 \right\rangle \quad (46)$$

XI.4.4. Ensemble o colectivo isobárico isotérmico (NPT)

Las condiciones del colectivo NPT se muestran en la Tabla 11.4.

XI.4.4.1. Presión del sistema

La presión del sistema debe llegar al equilibrio antes de realizar la simulación de MD. La presión se calcula inicialmente como tensor \mathbf{P} (ecuación 47a) a partir del tensor virial Ξ (Ecuación 47b).⁹³

$$\mathbf{P} = \frac{2}{V} \left[\left(\sum_{i=1}^N \mathbf{E}_{k,i} \right) - \Xi \right] \quad (47a)$$

$$\Xi = -\frac{1}{2} \sum_{i < j} (\mathbf{r}_{ij} \otimes \mathbf{F}_{ij}) \quad (47b)$$

Y el escalar presión es la traza del tensor \mathbf{P} , la suma de los elementos diagonales del tensor (traza del tensor).

Tabla 11.4. Condiciones de NPT

Variable	Valor o argumento
Integrador	<i>Leap-frog</i>
Número de pasos	5000
Paso de tiempo	2 fs = 0.002 ps
Tiempo de simulación	10 ps
Asignación de velocidades	Nula
Restricciones	Holonómicas
Presión de referencia (P_0)	1 bar
Compresibilidad isotérmica del agua	$4.5 \times 10^{-5} \text{ bar}^{-1}$

XI.4.5. Simulación de dinámica molecular (MD)

XI.4.5.1. Método de integración de Verlet

El teorema de Taylor (Ecuación 48) establece que si una función tiene derivadas continuas hasta la de orden $n+1$ en un intervalo dado, la función se puede aproximar con una serie de n términos y un residuo R_n asociado al error de la aproximación.⁹⁷

$$r(t) = \sum_{i=0}^n \frac{(\Delta t)^i}{i!} \frac{d^i}{dt^i} r(t_0) + R_n \quad (48)$$

La ecuación 48 es la serie de Taylor de r alrededor de t_0 .

En la simulación de MD, las ecuaciones de Newton son resueltas simultáneamente en pasos cortos de tiempo (típicamente comprendido entre 1 y 10 fs) mediante el algoritmo de Verlet, el cual emplea dos aproximaciones en series de Taylor truncadas al cuarto término (ecuaciones 49a y 49b) para las posiciones del paso nuevo ($t+\Delta t$) y el previo ($t-\Delta t$) respectivamente, y sumándolas se obtiene la nueva posición (Ecuación 49c). El sistema mantiene la presión y la temperatura en los intervalos requeridos, y las coordenadas atómicas son escritas en un archivo de salida cada cierto período. Las coordenadas, como función de t , representan la trayectoria del sistema.

$$r(t + \Delta t) = \sum_{i=0}^3 \frac{(\Delta t)^i}{i!} \frac{d^i}{dt^i} r(t_0) \quad (49a)$$

$$r(t - \Delta t) = \sum_{i=0}^3 \frac{(-\Delta t)^i}{i!} \frac{d^i}{dt^i} r(t_0) \quad (49b)$$

$$r(t + \Delta t) = 2r(t) + \frac{F}{m} \Delta t^2 - r(t - \Delta t) + R_n \quad (49c)$$

El algoritmo de Verlet es casi exacto, siendo el error de truncación de la serie del orden de $(\Delta t)^4$.⁷⁷

Si la simulación de MD se realiza considerando los enlaces como osciladores armónicos según el término $i=2$ de la serie de Taylor. A una oscilación simple con un grado de libertad, el eje internuclear, le corresponde una posición al equilibrio en la cual la energía sea mínima, la desviación de esta posición genera una fuerza (Ecuación 50) que retorna el enlace al punto de equilibrio. La energía potencial del enlace, considerado como un resorte según la Ley de Hooke (Ecuación 51a), y la energía cinética asociada (Ecuación 51b) generan la ecuación diferencial de movimiento (Ecuación 52) o lagrangiano del subsistema.

$$F = -\frac{dU}{dr} = \frac{k(\|\vec{r}\| - \|\vec{r}_{eq}\|)}{\|\vec{r}\|} \vec{r} \quad (50)$$

$$E_p = -\frac{k(\|\vec{r}\| - \|\vec{r}_{eq}\|)^2}{2} \quad (51a)$$

$$E_k = \frac{m\|v\|^2}{2} \quad (51b)$$

$$\frac{d^2}{dt^2}r + \frac{k}{m}r = r'' + \omega^2r = 0 \quad (52)$$

La cual tiene raíces complejas conjugadas, y la solución (Ecuación 53) denota que el enlace tiene un movimiento armónico.^{93,98}

$$r = C_1 \cos \omega t + C_2 \sin \omega t \quad (53)$$

La energía potencial rotacional (Ecuación 54) y la energía traslacional, que equivale a la energía cinética, definen la energía interna U del sistema (Ecuación 55).

$$E_{rot} = \frac{m \|\vec{r}\|^2 \omega^2}{2} \quad (54)$$

$$U = \sum_{i=1}^N E_{p,i} + \sum_{i=1}^N E_{k,i} \quad (55)$$

XI.4.5.2. Intervalo de tiempo de integración

El intervalo de tiempo de integración (τ) define la simulación correcta en la trayectoria. Con valores pequeños de τ la trayectoria de la simulación es más parecida al resultado de la integración analítica. Un valor típico de τ para simulaciones atomísticas es de 0.1 o 0.2 ps.

XI.4.5.3. Condiciones de la dinámica

Los colectivos o ensambles son las condiciones de los microestados del sistema. El colectivo canónico define constantes al número de partículas, el volumen y la temperatura (NVT), y el colectivo isobárico isotérmico (NPT) mantiene constantes el número de partículas, la presión y la temperatura. Al inicio de la simulación MD debe establecerse un estado inicial.

Las velocidades iniciales de los átomos se asignan aleatoriamente con una distribución de Maxwell-Boltzmann. La ecuación 30 representa la probabilidad de que el sistema se encuentre en el estado energético E_i .

XI.4.5.4. Temperatura y termostato de Berendsen

El algoritmo de Berendsen o termostato de Berendsen (Ecuación 56) previene desviaciones de la temperatura del sistema. Esta desviación decae exponencialmente con la constante τ .⁹³

$$\frac{dT}{dt} = \frac{T - T_0}{\tau}, \quad T = T_0 \left(1 - e^{-\frac{t}{\tau}}\right) \quad (56)$$

XI.4.5.5. Presión y barostato de Parrinello-Rahman

En el barostato de Parrinello-Rahman la matriz del vector de la caja de simulación B obedece otra relación de movimiento (Ecuación 57).

$$\frac{d^2 B}{dt^2} = VW^{-1}(B^T)^{-1}(P - P_{ref}) \quad (57)$$

Donde W es un parámetro matricial que es función de la compresibilidad isotérmica y la constante τ_P .

Con el barostato de Parrinello-Rahman se modifican las ecuaciones de movimiento (Ecuación 58). La constante τ_P debe ser mayor a la constante τ , ya que si el sistema se encuentra fuera del equilibrio resulta en oscilaciones muy grandes.⁹³

$$\frac{d^2 r}{dt^2} = \frac{F}{m} - M \frac{dr}{dt} \quad (58)$$

M es una matriz de modificación de las matrices B .

XI.4.5.6. Límites del sistema

Las condiciones periódicas en la frontera (*periodic boundary conditions, PBC*) simulan un sistema infinito, donde la celda de simulación está rodeada de réplicas de sí misma, evitando desviaciones de los átomos limítrofes respecto a los átomos centrales del sistema.

XI.4.5.7. Sumas de Ewald y PME

Está basado en la expresión culómbica de la energía potencial electrostática, y define las interacciones de las partículas en la celda de simulación con sus correspondientes partículas en *PBC*.

En una celda Ω con N cargas puntuales $\{q_N\}$ en posiciones respectivas $\{\vec{r}_N\}$, definida por tres vectores $\vec{e}_1, \vec{e}_2, \vec{e}_3$, la carga q_i en \vec{r}_i tiene una interacción culómbica con q_j en \vec{r}_j ($i \neq j$), así como con sus imágenes periódicas en $\vec{r}_j + n_1 \vec{e}_1 + n_2 \vec{e}_2 + n_3 \vec{e}_3$. La carga q_i también interactúa con sus propias imágenes periódicas en $\vec{r}_i + n_1 \vec{e}_1 + n_2 \vec{e}_2 + n_3 \vec{e}_3$.⁹⁹ La energía electrostática de Ω (ecuación 59) tiene la serie infinita exterior cuya convergencia depende de los valores de n_1, n_2 y n_3 .¹⁰⁰

$$E(r_1, \dots, r_N) = \frac{1}{2} \sum_{n \neq 0} \sum_i \sum_{j \neq i} \frac{q_i q_j}{\|r_i - r_j + n\|} \quad (59)$$

El método de sumas de Ewald (ecuación 60a) transformó la serie infinita por dos términos que convergen absolutamente, una suma directa en el espacio lineal aritmético \mathbb{R}^3 (E_{dir} , Ecuación 60b) y una suma recíproca en el espacio de Fourier (E_{rec} , Ecuación 60c). Las interacciones no enlazantes y no periódicas no se calculan y pertenecen a un tercer término de corrección (E_{cor}).

$$E = E_{dir} + E_{rec} + E_{cor} \quad (60a)$$

$$E_{dir} = \sum_n \frac{1 - \text{erf}(\beta \|(\vec{r} + n)\|)}{\|(\vec{r} + n)\|} \quad (60b)$$

$$E_{rec} = \frac{1}{\pi V} \sum_m \frac{1}{\|\vec{m}\|^2} e^{-\left(\frac{\pi\|\vec{m}\|}{\beta}\right)^2} e^{2\pi i \vec{m} \cdot \vec{r}} \quad (60c)$$

Donde $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ es la función error de x .

El primer sumando equivale físicamente a rodear cada carga puntual en el sistema por una distribución, típicamente normal, neutralizante, de cargas de igual magnitud y signo contrario. Éste término es responsable de las interacciones de corto alcance.

El segundo sumando compensa la distribución neutralizante del primer término, mediante una distribución compleja de cargas de signo opuesto a las del espacio real.⁷⁷ La base del espacio lineal de tres dimensiones son los vectores ortogonales de posición $\hat{i}, \hat{j}, \hat{k}$ ¹⁰¹ y la base del espacio de Fourier son los vectores de onda $m_1 = 2\pi \frac{j \times \hat{k}}{(i \times j) \cdot \hat{k}}$, $m_2 = 2\pi \frac{\hat{k} \times i}{(i \times j) \cdot \hat{k}}$ y $m_3 = 2\pi \frac{i \times j}{(i \times j) \cdot \hat{k}}$.¹⁰²

Las aproximaciones de la suma de Ewald, basadas en mallas (*particle mesh Ewald, PME*) aceleran la solución del método. En lugar de considerar una densidad continua de carga, se aproximan las cargas distribuidas en una malla⁷⁷ sobre la celda de simulación, discretizando los valores y aplicando la transformada rápida de Fourier (FFT, un método rápido de cálculo de la transformada discreta de Fourier). Se simula un cubo de 0.16 nm de lado en el espacio de Fourier para la suma recíproca.

XI.4.5.8. Interacciones de van der Waals

Estas interacciones pueden ser dipolo-dipolo o interacciones de Keesom, dipolo-dipolo inducido o interacciones de Debye, y dipolo instantáneo-dipolo inducido o interacciones de London.

Las interacciones de Keesom se consideran como de orientación fija o de media angular. Las interacciones de orientación fija tienen cuatro variables, la distancia de separación entre los dipolos, y tres ángulos y la energía de la interacción depende de la posición de los átomos en el sistema. La energía de interacción de media angular asume las fluctuaciones medias ponderadas de los dipolos en función de la temperatura.¹⁰³

La energía de las interacciones de Debye considera el campo eléctrico y una media ponderada del ángulo debido a la posición del dipolo.

La energía de las interacciones de London no incluye momentos dipolares, sino una función de las primeras energías de ionización de los átomos i,j involucrados.

Los tres tipos de energías de van der Waals se engloban aproximadamente en el término con exponente 6 en la ecuación de Lennard-Jones (Figura 11.4).¹⁰⁴

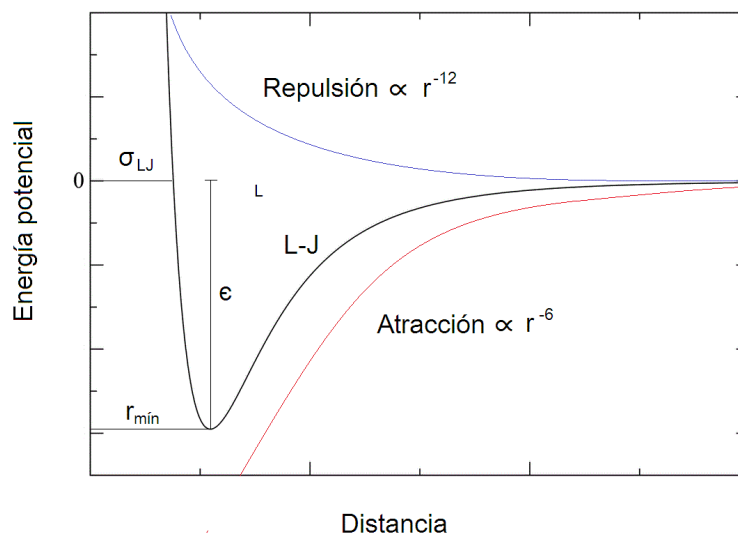


Figura 11.4. Gráfica de la energía potencial de atracción y repulsión, y su aproximación por la función de Lennard-Jones (L-J).

XI.4.5.9. Interacciones hidrofóbicas

La atracción hidrofóbica es la fuerza no covalente y no electrostática más importante entre partículas apolares. En un medio acuoso, la atracción hidrofóbica se debe a la energía libre de los enlaces de hidrógeno entre las moléculas de agua del entorno, siendo la energía de la atracción hidrofóbica igual a la energía libre de cohesión en el agua.¹⁰⁵

XI.4.6. Análisis de MD

XI.4.6.1. Agrupamiento o *clustering*

Es una técnica empleada para seleccionar un estado representativo de un conjunto heterogéneo de conformaciones de la macromolécula que ha sido simulada por MD. En el *clustering*, las conformaciones se agrupan de manera que, aquellas pertenecientes a un mismo grupo o *cluster*, compartan un alto grado de similitud entre sí y sean distantes de las situadas en otros *clusters*.⁷⁷ El modo de agrupamiento se define por un rango o *cutoff* de RMSD, el cual define el valor máximo de desviación para cada *cluster*. En cada uno, se calculan coordenadas medias para generar una conformación representativa. Se puede generar una matriz con los valores de RMSD a través del

tiempo, representado con colores correspondientes a su magnitud. A esta representación se le conoce como matriz de RMSD.

XI.4.6.2. Raíz de la desviación cuadrática media (*Root-Mean-Square Deviation*)

El análisis de los resultados de las simulaciones de MD se basa en las desviaciones de la posición del átomo i -ésimo respecto a las coordenadas iniciales a través del tiempo de simulación. El valor de RMSD (Ecuación 61) expresa la magnitud media de esta desviación. Valores de RMSD mayores a 2.0 Å indican un desplazamiento significativo.

$$RMSD(t) = \langle (\Delta r_i)^2 \rangle = \sqrt{\frac{\sum_{i=1}^N m_i \|r_i(t) - r_i(t_0)\|^2}{\sum_{i=1}^N m_i}} \quad (61)$$

(a) Equilibrios del sistema

Se considera que el sistema se encuentra en equilibrio mecánico cuando sus valores de RMSD tienen oscilaciones mínimas como función del tiempo; cuando la amplitud de la función periódica tiende a cero. Debido a la naturaleza de las gráficas de valores de RMSD en función del tiempo, se puede elegir el intervalo de tiempo aproximado al cual se considera el sistema equilibrado, o puede aplicarse a la función una transformación en la cual se filtre o suavice la función y sea más simple la identificación del intervalo de equilibrio. Un método de filtrado de la función RMSD es la transformada discreta de Fourier.

Una serie de Fourier es la representación de una función f como una serie de constantes multiplicadas por las funciones seno y coseno de diferentes frecuencias. Si la función $f(t)$ está definida en un intervalo $[-L, L]$, la serie de Fourier de $f(t)$ (Ecuación 62) puede aproximar cada valor de $f(t)$ ¹⁰¹. Las series de Fourier se pueden representar con funciones seno y coseno o con la función exponencial compleja (Ecuación 63),^{34,106-108} cuya transformación se basa en la fórmula de Euler.

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{n\pi x}{L}\right) + b_n \text{sen}\left(\frac{n\pi x}{L}\right) \right] \quad (62)$$

$$f(t) = \sum_{n=-\infty}^{\infty} d_n e^{in\omega x} \quad (63)$$

Donde $\omega = \frac{2\pi}{p}$ es la frecuencia, y a_0 , a_n , b_n y d_n son coeficientes de Fourier.

La transformada de Fourier de f (Ecuación 64) convierte a $f(t)$ en $\mathcal{F}\{f(t)\}$. Si M es un enteros positivo y $\{u_j\}_{j=0}^{N-1}$ es una sucesión de N números complejos, la transformada discreta de Fourier de N puntos de

u es la serie DFT{u} (Ecuación 65). Obteniendo la parte real de $\mathcal{F}\{f(t)\}$ y calculando la transformada inversa discreta de Fourier (Ecuación 66) de ésta, se obtiene la función f(t) filtrada o suavizada. Frecuentemente este análisis se realiza para eliminar ruido e interferencias de una función temporal.¹⁰⁸

$$\mathcal{F}\{f(t)\}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \quad (64)$$

$$DFT\{u\}(r) = \sum_{j=0}^{N-1} u_j e^{-\frac{2\pi i j r}{N}} \quad (65)$$

$$DFT^{-1}\{DFT[f(t)]\}(t) = \frac{1}{N} \sum_{r=0}^{N-1} DFT\{u\}(r) e^{\frac{2\pi i j r}{N}} \quad (66)$$

XI.4.6.3. Raíz de la fluctuación cuadrática media (*Root-Mean-Square Fluctuation*)

El gráfico de RMSF (Ecuación 67) en función del tiempo representa las desviaciones de posición por residuo de aminoácido.¹⁰⁹

$$RMSF(t) = \sqrt{\frac{\sum_t^{n_t} \langle [r_i(t) - r_i(t_0)]^2 \rangle}{n_t}} \quad (67)$$

Donde la suma se hace en la media de la distancia cuadrática de los átomos que contiene la cadena lateral de cada residuo de aminoácido, desde el paso de tiempo inicial ($t = 0$, o según se requiera analizar), hasta el último paso de tiempo (n_t). El análisis de RMSF suele realizarse desde el tiempo t en que el sistema ha alcanzado el equilibrio hasta el final de la simulación para evitar fluctuaciones de gran magnitud. Dependiendo del sistema, este valor de t puede variar considerablemente, y debe encontrarse de forma particular.

XI.4.6.4. Radio de giro

El radio de giro de una proteína (Ecuación 68) es la distancia del i -ésimo átomo al centro de masa. El centro de masa (Ecuación 69) es el punto donde las posiciones ponderadas relativas de los componentes suman cero. Aumentos progresivos o bruscos del radio de giro a través del tiempo indican disociación de subunidades o pérdida de estructuras secundarias y terciarias.

$$rg(t) = \sqrt{\frac{\sum_{i=1}^N [m_i (\|\vec{r}_i - \vec{r}_{bar}\|)^2]}{\sum_{i=1}^N m_i}} \quad (68)$$

$$\vec{r}_{bar} = \frac{\sum_{i=1}^N m_i \vec{r}_i}{\sum_{i=1}^N m_i} \quad (69)$$

XI.4.6.5. Puentes de hidrógeno

Un puente de hidrógeno existe cuando un átomo de hidrógeno enlazado a un átomo electronegativo interactúa con otro átomo electronegativo.¹¹⁰ El puente de hidrógeno se forma a distancias menores que las calculadas con base en los radios de van der Waals. En los átomos A...H-B, la nube electrónica del átomo B experimenta compresión por el hidrógeno, lo que indica que está más próximo a éste. Típicamente el enlace de hidrógeno se forma entre dos átomos electronegativos, donde el enlace con el átomo B es covalente, con orden de enlace cercano a uno, y el “enlace” con el átomo A es más débil y más largo, con una entalpía de enlace de 50 kJ/mol o menor, y una distancia media de 2 Å.¹¹⁰ Cuando dos átomos electronegativos A,B, uno de ellos enlazado a H, se sitúan a una distancia de aproximadamente 3.0 Å, el átomo de hidrógeno puede formar un “enlace” con el otro átomo electronegativo. Durante la simulación MD, el cambio de posición atómica respecto al tiempo puede formar y romper enlaces de hidrógeno entre los átomos electronegativos. Cuando dos subestructuras forman múltiples puente de hidrógeno, el trabajo necesario para aumentar la distancia entre ellas es mayor, y pueden permanecer unidas durante un intervalo de tiempo en la simulación. Para analizar la formación de puentes de hidrógeno en los sistemas a través del tiempo, se cuentan los grupos de átomos A, H, B, donde la distancia de van der Waals entre A y B sea menor a 3.4 Å.

Si se tiene un conjunto de sistemas donde se forman y rompen puentes de hidrógeno en la simulación MD, se puede comparar el conteo de enlaces formados en el tiempo de simulación con una prueba estadística no paramétrica para saber si alguno de los sistemas tiene mayor significancia en la formación de estos enlaces, como lo es la prueba de Mann-Whitney.

(a) Prueba de Mann-Whitney.

Esta prueba es la alternativa no paramétrica, o de libre distribución, a la prueba t de comparación de dos medias. Esta comparación prueba la hipótesis nula de que las medianas de los sistemas muestreados son estadísticamente iguales. La prueba consiste en el ordenamiento de los valores de menor a mayor y la asignación de un rango a cada uno. Se calcula la suma de los rangos de cada muestra para establecer el estadístico de prueba U (Ecuación 70). Los estadísticos U_i y U_{i+1} son valores de variables aleatorias que tienen media μ (Ecuación 71) y varianza σ^2 (Ecuación 72).¹¹¹ Se calcula un valor z_{rk} (Ecuación 73) que se contrasta con el valor z_c de la distribución normal.

$$U_i = \sum_j rk_j - \frac{n_i(n_i + 1)}{2} \quad (70)$$

$$\mu = \frac{n_i n_{i+1}}{2} \quad (71)$$

$$\sigma^2 = \frac{n_i n_{i+1} (n_i + n_{i+1} + 1)}{12} \quad (72)$$

$$z_{rk} = \frac{U_i - \mu}{\sigma} \quad (73)$$

XI.4.6.6. Análisis de estructura secundaria

La estructura secundaria de una proteína se refiere a disposiciones particularmente estables de los aminoácidos que dan lugar a patrones estructurales repetitivos. El análisis de estructuras secundarias a través del tiempo de simulación consiste en el cálculo de los ángulos diedros ϕ y ψ de la cadena principal de la proteína.

XI.4.6.7. Matrices de distancias

La matriz de distancias mínimas entre dos residuos de aminoácidos A_i , A_j en una proteína se define como la distancia más pequeña entre los pares de átomos $i \in A_i$, $j \in A_j$. En la gráfica de la matriz se pueden encontrar formación de puentes salinos.⁹³

XI.4.6.8. Energía de unión

(a) Área superficial de Poisson-Boltzmann (PBSA)

El método de mecánica molecular-área superficial de Poisson-Boltzmann (MM/PBSA) es uno de los enfoques más utilizados para la estimación de energías de unión en biomoléculas.⁵⁰ MM/PBSA incluye tres términos energéticos: el cambio del potencial en el vacío, el cambio del potencial de solvatación y un término entrópico. El primer término E_{MM} (Ecuación 74) es la energía potencial de mecánica molecular que considera las interacciones enlazantes y no enlazantes. Las interacciones enlazantes consisten en las interacciones de enlace, angulares, diédricas e impropias. Las interacciones no enlazantes engloban las interacciones coulombicas y de Lennard-Jones.

$$E_{MM} = E_{bond} + E_{angle} + E_{dihed} + E_{imp} + E_{elec} + E_{L-J} \quad (74)$$

El segundo término G_{solv} , energía de solvatación (Ecuación 75), es la contribución requerida para transferir el soluto del vacío al solvente, y se compone de los términos electrostático (G_{pol}) y no electrostático (G_{apol}). El término electrostático se estima resolviendo la ecuación de Poisson-Boltzmann (Ecuación 76a de manera general y 76b según la teoría de Debye-Hückel). La ecuación 78a es una ecuación diferencial parcial no lineal, que incluye el potencial dependiente de la posición $\Phi(r)$ como derivada parcial de segundo orden y como argumento de una exponencial; y la constante dieléctrica ϵ . Su resolución implica un elevado tiempo computacional. En la ecuación 76b, una

simplificación, el término κ^2 (Ecuación 76c) depende de la fuerza iónica de la solución. El término no electrostático abarca el trabajo realizado por el soluto para formar una cavidad en el solvente, y un término de van der Waals.

$$G_{solv} = G_{pol} + G_{cav} + G_{vdW} \quad (75)$$

$$\nabla^2 \Phi(r) = -\frac{F}{\epsilon\epsilon_0} \sum_i z_i C_i e^{-\frac{z_i e \Phi(r)}{k_B T}} \quad (76a)$$

$$\nabla^2 \Phi(r) = \kappa^2 \Phi(r) \quad (76b)$$

$$\kappa = \frac{2z_i^2 e F C_i}{k_B T \epsilon \epsilon_0} \quad (76c)$$

MM/PBSA también descompone la energía de unión en contribuciones individuales por residuo de aminoácido y el ligando.⁵⁰ La entropía de unión del sistema proteína-ligando no se incluye en MM/PBSA.

Las condiciones del análisis de MMPBSA (Tabla 11.5) pueden simplificar el cálculo de $\Phi(r)$ de manera considerable.

Tabla 11.5. Condiciones del método MMPBSA

Variable	Valor o argumento
Carga de cationes (z)	1
Radio de cationes	0.95 nm
Carga de aniones	-1
Radio de aniones	1.81
Constante dieléctrica del soluto	2
Constante dieléctrica del solvente	80
Temperatura de referencia (T_0)	300 K
Energía superficial (γ_{sol})	0.0226778 kJ/molÅ ²
Condiciones periódicas	Debye-Hückel múltiple
Método de solución de PB	Ecuación lineal de Poisson-Boltzmann

Nota: El modelo de Debye-Hückel expande la exponencial con una serie de Taylor, linealizando la ecuación 76a.

(b) Aproximación lineal a la energía de unión (LIE)

Para un sistema con dos ligandos A y B, con sus respectivos potenciales V_A y V_B (Figura 11.5), en un ciclo termodinámico. De éste ciclo se obtiene que $\Delta G_{\text{bind},B} - \Delta G_{\text{bind},A} = \Delta^2 G_{\text{solv,prot}} - \Delta^2 G_{\text{solv,H}_2\text{O}}$. Considerando que $\Delta G_{\text{bind},A}=0$, la función de energía libre (Ecuación 77) está relacionada con la media de la densidad de probabilidad de una coordenada en el potencial superficial V .¹¹²

$$\Delta G_{BA} = -k_B T \ln \left\langle e^{-\frac{V_A - V_B}{k_B T}} \right\rangle \quad (77)$$

Realizando la expansión con series de potencias para el logaritmo y la exponencial (Ecuación 78), se puede aproximar ΔG considerando las contribuciones culómbicas y de van der Waals, con el primer término de la respectiva serie (Ecuación 79) que asume una dependencia lineal.

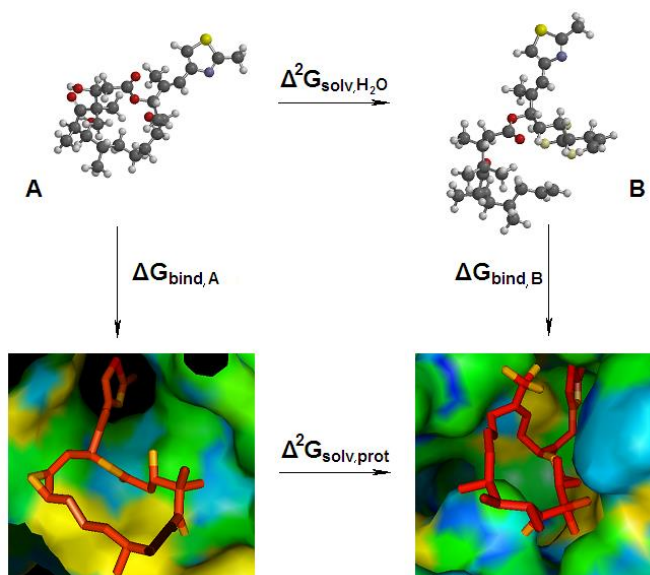


Figura 11.5. Ciclo termodinámico en el que se basa la función de energía libre.

$$\Delta G_{BA} = -k_B T \sum_{m=1}^{\infty} \frac{1}{m} \sum_{n=0}^{\infty} \frac{\left(-\frac{V_B - V_A}{k_B T}\right)^{nm}}{n!^m} \quad (78)$$

$$\Delta^2 G = \frac{1}{2} \langle \Delta V_{elec} \rangle + c \langle \Delta V_{vdW} \rangle \quad (79)$$

XII. ANEXO III

XII.1. Generación de todos los modelos lineales posibles

Para evaluar la significancia global de todos los modelos posibles con uno a diecinueve descriptores generados en *FAFDrugs3*, se usó la suma de combinaciones (Ecuación 80) en el lenguaje Bash.

$$\sum_{x=1}^{19} C_x^{19} \quad (80)$$

Una vez generados los 524,287 modelos, se aplicó una regresión lineal múltiple a cada uno de ellos con el lenguaje R. Un resumen de las instrucciones más fundamentales se muestra en la Tabla 12.1.

Tabla 12.1. Instrucciones básicas para el análisis preliminar de regresión múltiple

Instrucción	Acción
<code>head -n1 \$CSV sed 's/,/ /g' > List_descriptors.erase</code>	Agrupar los nombres de los descriptores del archivo de entrada *.csv que contiene la actividad experimental y los valores de los descriptores
<code>read -p "Observable>>" ACT</code>	Definición de la actividad o propiedad biológica
<code>descriptors=\$(cut -d ' ' -f2- List_descriptors.erase cut -d ' ' -f2-)</code>	Definición del conjunto de descriptores en un archive de salida
<code>combine \$descriptors sed 's/ / + /g' >> all.models.dat</code>	Agrupamiento de los descriptores, un modelo por renglón
<code>echo "data <- read.csv("\$CSV")" >> mlr.r</code>	Ingreso de la variable *.csv que contiene los descriptores de cada modelo
<code>echo "model <- lm(\$ACT ~ \$descriptors, data=data)" >> mlr.r</code>	Aplicación de la función " <i>linear model: lm</i> " con la variable ACT (<i>activity: logBB</i>) y las variables regresoras <i>descriptors</i>
<code>echo "summary(model)" >> mlr</code>	Imprime el resumen de la regresión lineal múltiple: r^2 , \bar{r}^2 , F, p y grados de libertad; y la significancia del conjunto de descriptores, evaluados mediante una prueba t
<code>R -f mlr.r -slave > models.dat</code>	Ejecuta el análisis llamando al archivo de instrucciones "mlr.r" y guarda el resultado en el archive de salida "models.dat"

XII.2. Residuales estudentizados de los seis modelos lineales más significativos

Los residuales estudentizados para cada uno de los seis modelos lineales más significativos (Figura 12.1) y su tipo de densidad (Tabla 12.2) se analizaron para elegir un criterio adecuado para la detección de valores atípicos.

Para el modelo de dos descriptores, no se puede rechazar H_0 de normalidad. Los estadísticos de los residuales fueron $\bar{e} = -1.79 \times 10^{-17}$ y $s_e^2 = 0.067$. Los residuales estudentizados tuvieron $\bar{r} = -0.0014$ y $s_r^2 = 1.0161$. La desviación estándar es 1.008, que es congruente con los supuestos de los residuales en regresión lineal,⁶³ y el gráfico de cuantiles normales (Figura 12.2) cualitativamente tiene una tendencia hacia la normalidad, con lo que se justificó la elección de los valores atípicos como aquellos fuera del intervalo $[-2,2]$ o aproximadamente el 95% de los valores, asumiendo los parámetros $\mu \approx 0$ y $\sigma \approx 1$. Para este mismo modelo, la dispersión de los valores predichos en función de los observados para logBB (Figura 12.3) también tiene correlación significativa ($r = 0.789$, $p < 2.2 \times 10^{-16}$).

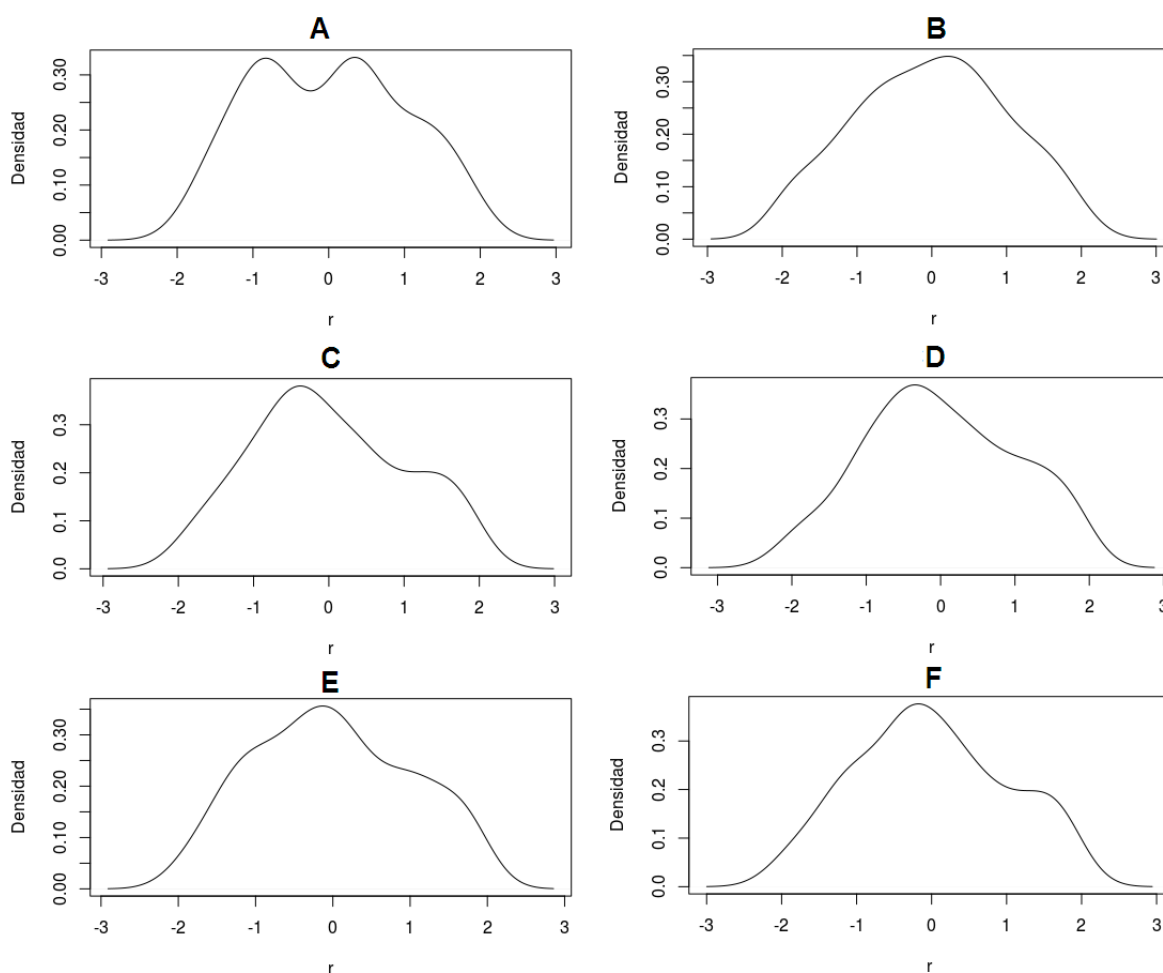


Figura 12.1. Densidades de los residuales estudentizados utilizados para la detección de valores atípicos de los modelos lineales QSPR con (A) uno (B) dos, (C) tres, (D) cuatro (E) cinco y (F) seis descriptores.

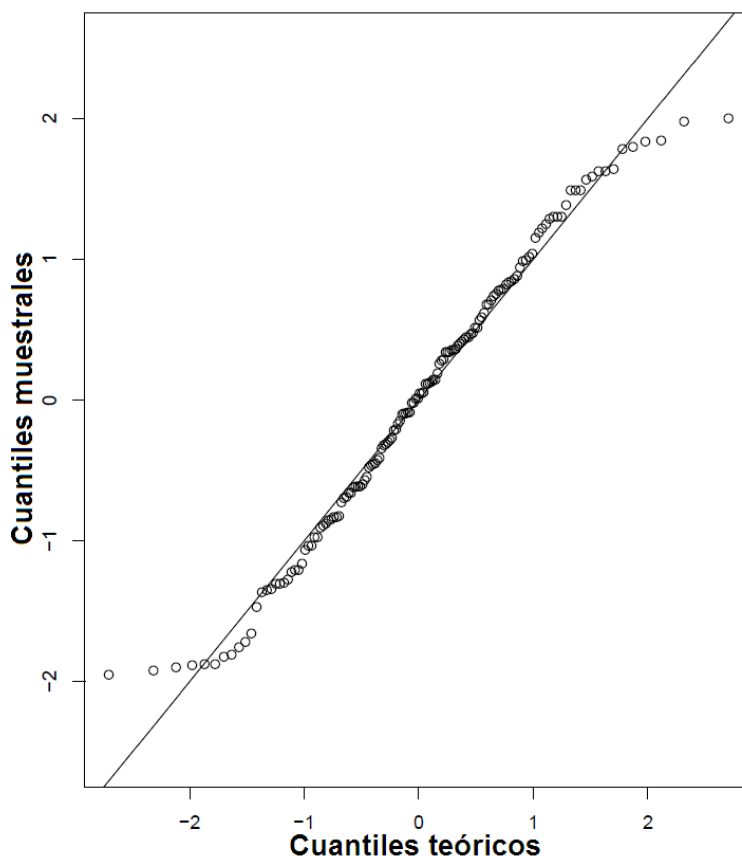


Figura 12.2. Gráfico de cuantiles normales para los residuales estudentizados para el modelo con dos descriptores. Los cuantiles muestrales son los calculados a partir de los residuales estudentizados del modelo y los teóricos corresponden a los cuantiles de la densidad de probabilidad normal.

Tabla 12.2. Características de los residuales estandarizados y estudentizados de los modelos lineales QSPR. La comparación de los residuales estandarizados y los residuales estudentizados no tuvo diferencia significativa ($\alpha=0.05$).

Descriptores en los modelos QSPR	Valor p de la comparación de d_i y r_i	Valor p de Shapiro-Wilk de r_i	Hipótesis de normalidad
1	0.9679	0.001415	Rechazada
2	0.959	0.03886	No se rechaza
3	0.9611	0.0037	Rechazada
4	0.9595	0.01435	Rechazada
5	0.9632	0.002765	Rechazada
6	0.9588	0.008187	Rechazada

Nota: Sólo el modelo lineal con dos descriptores, $\log BB = f(\log P + tPSA)$, generó un conjunto de residuales estudentizados con población normal, probada con la prueba de Shapiro-Wilk ($\alpha=0.05$).

Los residuales del modelo final de dos descriptores, al que se le eliminaron los valores atípicos mediante los residuales estudentizados del modelo inicial, tuvieron una densidad simétrica (Figura 12.4).

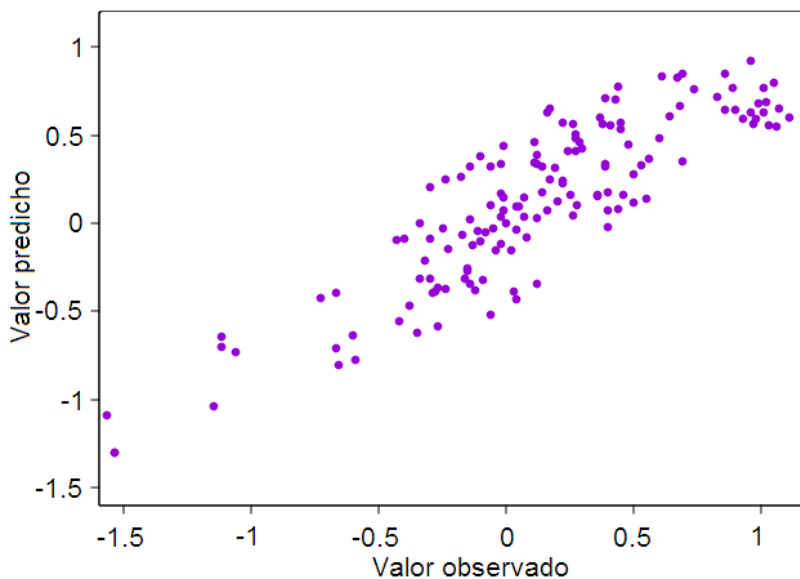


Figura 12.3. Gráfico de valor predicho contra el valor observado (experimental) de logBB para el modelo con dos descriptores. A partir de esta dispersión se calculan los estadísticos de validación r , r_{int}^2 , r_0^2 y $r_0'^2$.

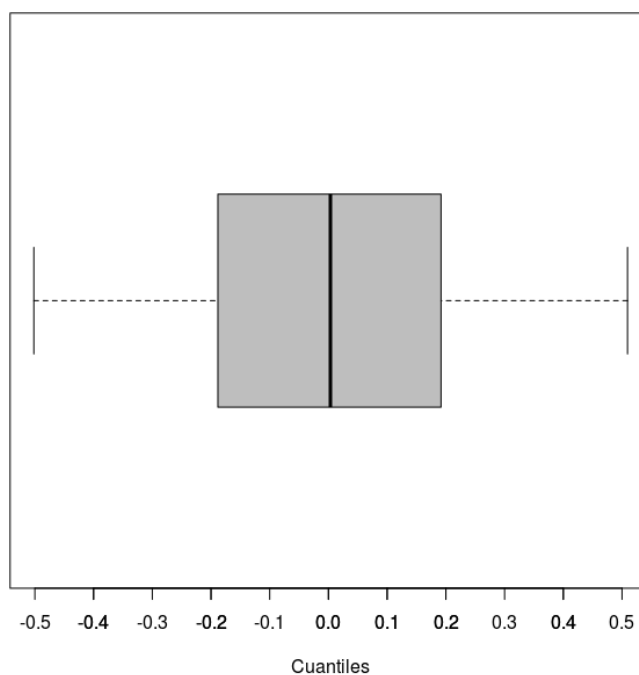


Figura 12.4. Gráfico de caja para los residuales del modelo final con dos descriptores. Los valores notables son: mín=-0.502, 1Q=-0.186, mediana=0.008, 3Q=0.195, máx=0.509.

XII.3. Simulación de MD clásica

En la simulación de MD clásica se modelan los átomos e iones como masas y cargas puntuales, pudiendo tener carga formal o parcial. Los enlaces covalentes se modelan como osciladores armónicos, y los enlaces salinos y de hidrógeno se determinan mediante los valores de energía electrostática coulombica, y de atracción o repulsión mediante Lennard-Jones dependientes de las distancias entre las partículas. Para el estudio de las interacciones que forman los estabilizadores microtubulares con los sistemas de tubulinas, los efectos electrónicos no son fundamentales.

XII.4. Ligandos estabilizadores acoplados en el heterodímero $\alpha\beta$

Se acoplaron tres ligandos estabilizadores en el heterodímero $\alpha\beta$, construido conforme a la estructura PDB: 1JFF. Dos ligandos se acoplaron en el sitio de los taxanos utilizando AutoDock 4.2 (LAU y DIC) y AutoDock Vina (DIC), y POU en un sitio adyacente utilizando AutoDock 4.2 conforme la estructura (Figuras 12.5 a 12.8).

En el sistema LAU- $\alpha\beta$ el sistema no alcanzó el equilibrio dinámico, aunque su función RMSD(t) tuvo valores máximos de menor magnitud que el sistema $\alpha\beta$. El ligando LAU se desacopló totalmente del dímero, solvatándose en la celda de simulación, para posicionarse en otro sitio del dímero aproximadamente a los 35 ns, donde permaneció hasta el final de la simulación. La solvatación de LAU, con XLOGP3 = 3.2, resultó más favorecida que su acoplamiento en el sitio de unión de los taxanos, lo que indica que no se acopló en el sitio preciso, adicionalmente a la inestabilidad del sistema dimérico. Al no alcanzar el equilibrio dinámico, las fluctuaciones de las cadenas laterales son grandes, aunque la integridad estructural de las subunidades no se modificó.

El sistema POU- $\alpha\beta$, mostró una inestabilidad mayor que el sistema con LAU, aunque se sigue notando una gráfica de RMSD(t) similar. El ligando POU se solvató y se separó de la proteína $\alpha\beta$. Al observar la estructura de POU (Ver sección II, Figura 2.5), los grupos hidroxilo y metoxilo favorecen la formación de puentes de hidrógeno con el agua, proceso que mostró ser más favorecido a comparación de la formación del complejo.

Para los sistemas DIC- $\alpha\beta$ se acopló el ligando mediante dos SF, AutoDock 4.2 y AutoDock Vina para realizar una comparación. Los acoplamientos fueron similares. Para la simulación con el acoplamiento mediante la SF semiempírica de AutoDock 4.2 las cadenas principales mostraron la tendencia de los sistemas anteriores, y en este caso DIC experimentó un desacoplamiento del sitio de unión, que incrementó los valores de la función RMSD(t) entre los 10 ns y los 30 ns de simulación. Por la alta hidrofobicidad del ligando (Ver sección II, Figura 2.5), no se espera que su solvatación esté

favorecida sobre la formación del complejo, por lo que regresó al sitio en los últimos 20 ns en una orientación diferente a la acoplada inicialmente. Los valores de RMSF disminuyeron respecto a los sistemas anteriores, y las estructuras terciarias se conservaron, comportamiento que empíricamente ha sido visto en los sistemas de tubulinas. El sistema en el que DIC se acopló con la SF rápida de AutoDock Vina tuvo tendencias de RMSD y rg de los polipéptidos similares al otro sistema de DIC, con mayores fluctuaciones de las cadenas laterales. El ligando presentó rotaciones múltiples en el sitio de unión, y de manera análoga, su solvatación no resultó estar favorecida.

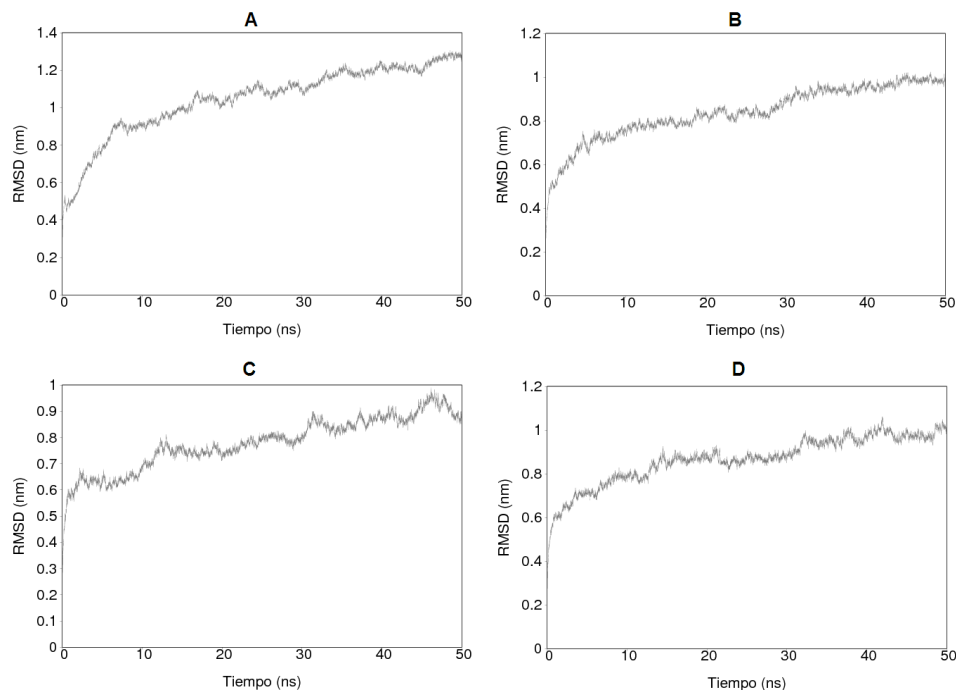


Figura 12.5. Gráficos de RMSD de las cadenas principales de los polipéptidos en función del tiempo para los sistemas heterodiméricos $\alpha\beta$ con los ligandos (A) LAU, (B) POU, (C) DIC acoplado con AutoDock 4.2 y (D) DIC acoplado con AutoDock Vina.

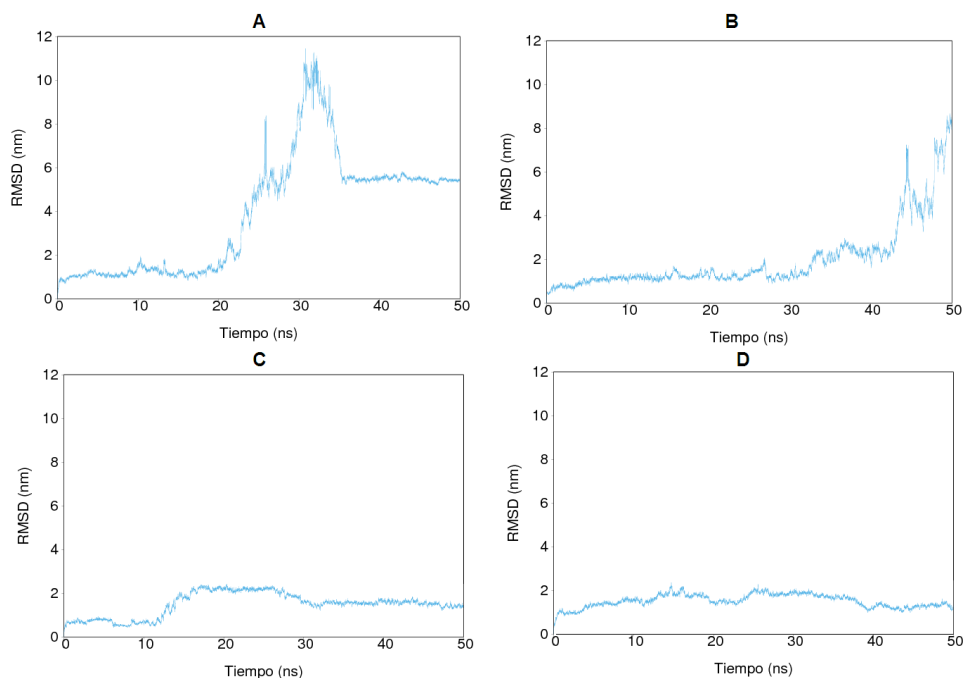


Figura 12.6. Gráficos de RMSD de los ligandos en el sitio de acoplamiento en función del tiempo. (A) LAU, (B) POU, (C) DIC acoplada con AutoDock 4.2 y (D) DIC acoplado con AutoDock Vina.

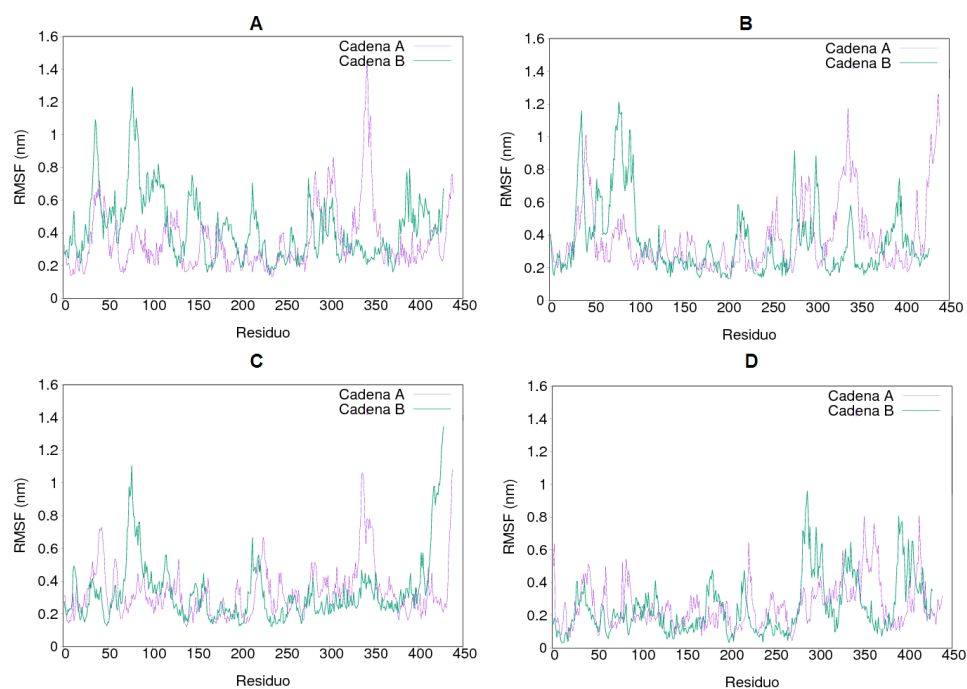


Figura 12.7. Gráficos de RMSF de los polipéptidos en función del tiempo para los sistemas heterodiméricos $\alpha\beta$ con los ligandos (A) LAU, (B) POU, (C) DIC acoplado con AutoDock 4.2 y (D) DIC acoplado con AutoDock Vina.

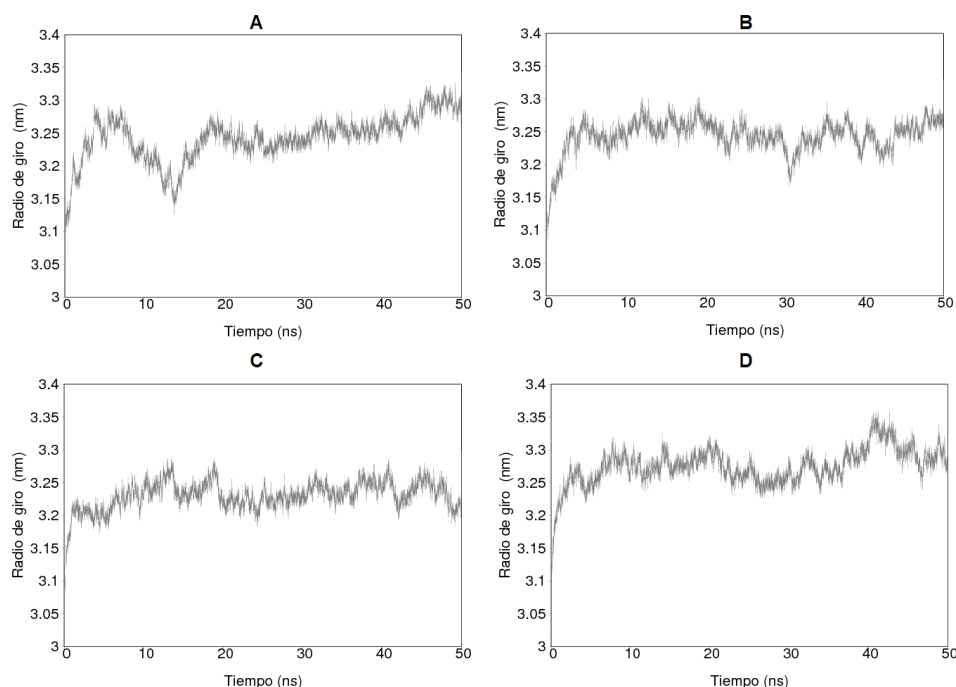


Figura 12.8. Gráficos de radio de giro de los polipéptidos en función del tiempo para los sistemas heterodiméricos $\alpha\beta$ con los ligandos (A) LAU, (B) POU, (C) DIC acoplado con AutoDock 4.2 y (D) DIC acoplado con AutoDock Vina.

La escasa afinidad de los ligandos con el heterodímero $\alpha\beta$ tiene relación con la baja espontaneidad de la unión a dímeros aislados en una célula, y su impacto casi nulo en células que no se encuentran en mitosis, meiosis y procesos hiperactivos de tráfico vesicular o quimiotaxias,²⁰ como los que presentan las células de linaje hematopoyético.

XII.5 RMSD en el sistema $\alpha_2\beta_2$

El análisis de agrupamiento del sistema $\alpha_2\beta_2$ con rango de 0.09 nm para la cavidad de EPA tiene una densidad de RMSD (Figura 12.9) considerando las conformaciones con distancia mayor a 0.9 nm entre los centros de masa de los residuos Q279 y L361 utilizando MDAnalysis.⁴⁸ Al aplicar la prueba de Shapiro-Wilk, la hipótesis de que la muestra proviene de una población normal se rechaza al 95% de confianza ($p = 0.0016$). La matriz de RMSD (Figura 12.10) para la cavidad tiene un intervalo de 0.054 nm a 0.194 nm con media de 0.12 nm.

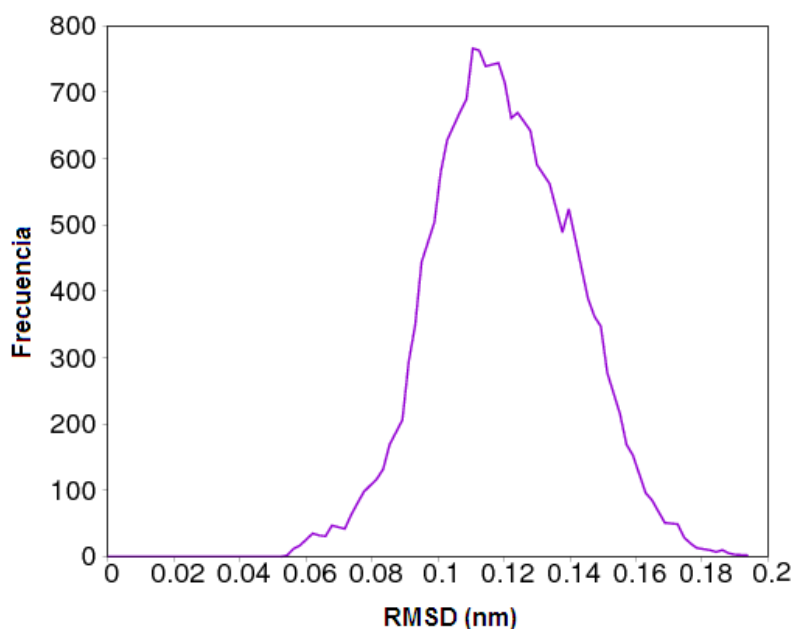


Figura 12.9. Polígono de frecuencia del análisis de agrupamiento.

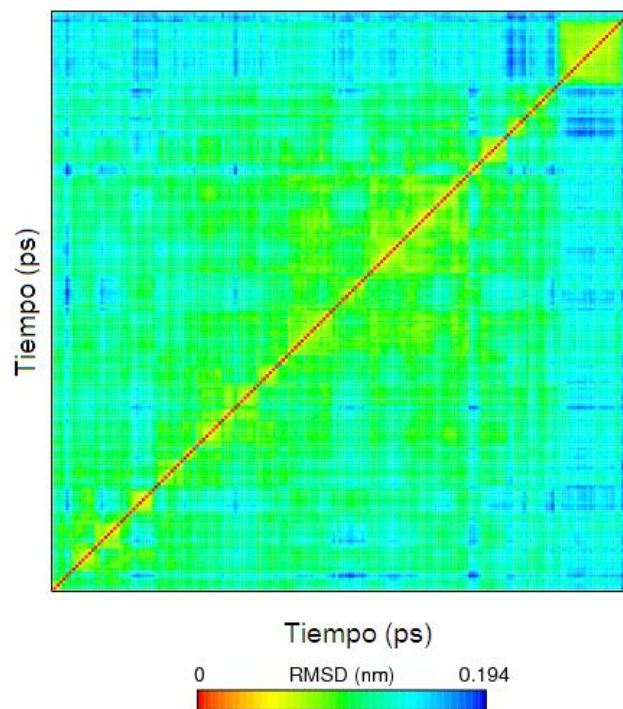


Figura 12.10. Matriz de RMSD del sitio de unión de EPA con distancia mínima de 0.9 nm entre los centros de masa de Q279 y L361.

Del análisis de agrupamiento de conformaciones, se eligieron los criterios empíricos de un número de grupos que tuvieran una relación tipo exponencial con su tamaño (Figura 12.11), seleccionándose así el rango o *cutoff* de 0.09 nm.

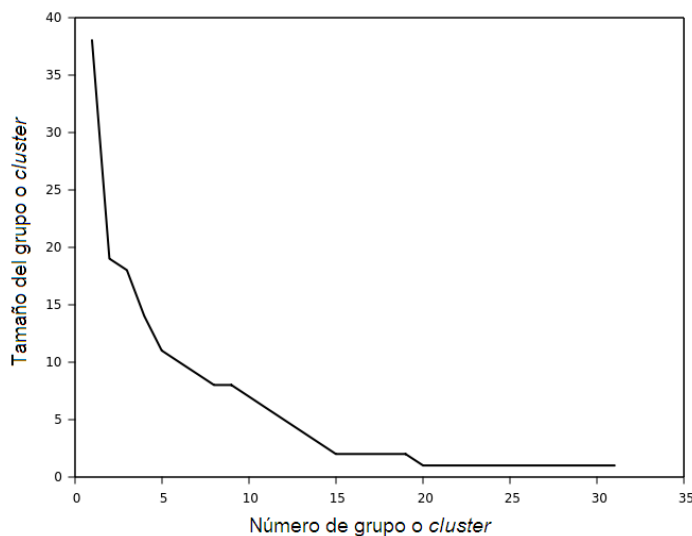


Figura 12.11. Relación del tamaño de *cluster* con su abundancia. La relación no debe ser lineal.

XII.6. Sistema homodimérico EPA₂-β₂

XII.6.1. Variables termodinámicas

Las gráficas de las variables E_{pot} , T , P y V a través del tiempo para el control positivo se muestran en la Figura 12.12. En EM se alcanzó el equilibrio mecánico, la función de E_{pot} fue decreciente y asintótica en el valor mínimo. En el colectivo NVT, la temperatura se mantuvo constante. En el ensamble NPT la presión del sistema tuvo las mayores variaciones al principio de la simulación NPT, que disminuyeron hacia los 2 ps. Las fluctuaciones de la presión se deben al comportamiento de las moléculas de agua, con un índice de compresibilidad definido, y las restricciones de los polipéptidos que ejercen una presión de oposición al solvente. En la simulación de MD, la presión tuvo una tendencia similar a la del ensamble NPT, sin que la fuerza superficial en el sistema llegara a desnaturalizar a la proteína. La temperatura y el volumen del sistema se mantuvieron con variaciones muy bajas, con rangos de 4 K y 16 nm³ aproximadamente.

La energía potencial de los diez sistemas homodiméricos se minimizó en la simulación de EM al optimizar la geometría. Durante la simulación los sistemas se deben mantener en condiciones isotérmicas, isobáricas e isocóricas. Durante la simulación NVT, la temperatura de referencia fijada en 300 K tuvo oscilaciones de amplitud muy baja. La presión de referencia, fijada en 1 bar, tuvo oscilaciones de gran amplitud durante la simulación de NPT. La variación de la presión responde a la variación de las contribuciones de la energía cinética, el tensor virial y la energía superficial durante el ensamble, ya que el solvente posee un coeficiente de compresibilidad isotérmica definido y el sistema proteico tiene restricciones en la cadena principal, provocando incremento de la energía cinética en el solvente y consecuentemente un aumento de la tensión superficial del sistema proteico sometido a restricciones. Si el cambio de presión respecto al tiempo llega a ser suficiente, los polipéptidos

pueden desnaturalizarse, fenómeno que no se presentó al observar los gráficos de radio de giro. De manera general, el control positivo presentó oscilaciones en la presión de menor amplitud que el control negativo. Durante la simulación MD, la temperatura y el volumen tuvieron poca variación. La temperatura del sistema es corregida mediante el termostato de Berendsen, el cual provoca el decaimiento exponencial de la desviación de la temperatura,⁹³ que la sitúa aproximadamente en $300 \text{ K} \pm 2 \text{ K}$. El volumen tuvo un intervalo de variación de no más de veinte nm^3 , el cual es dependiente de la variación de la presión. La presión del sistema manifestó nuevamente oscilaciones muy grandes, que se presentan por la redefinición de las ecuaciones de movimiento, aunque el barostato de Parrinello-Rahman limita la tendencia errática del escalar presión.

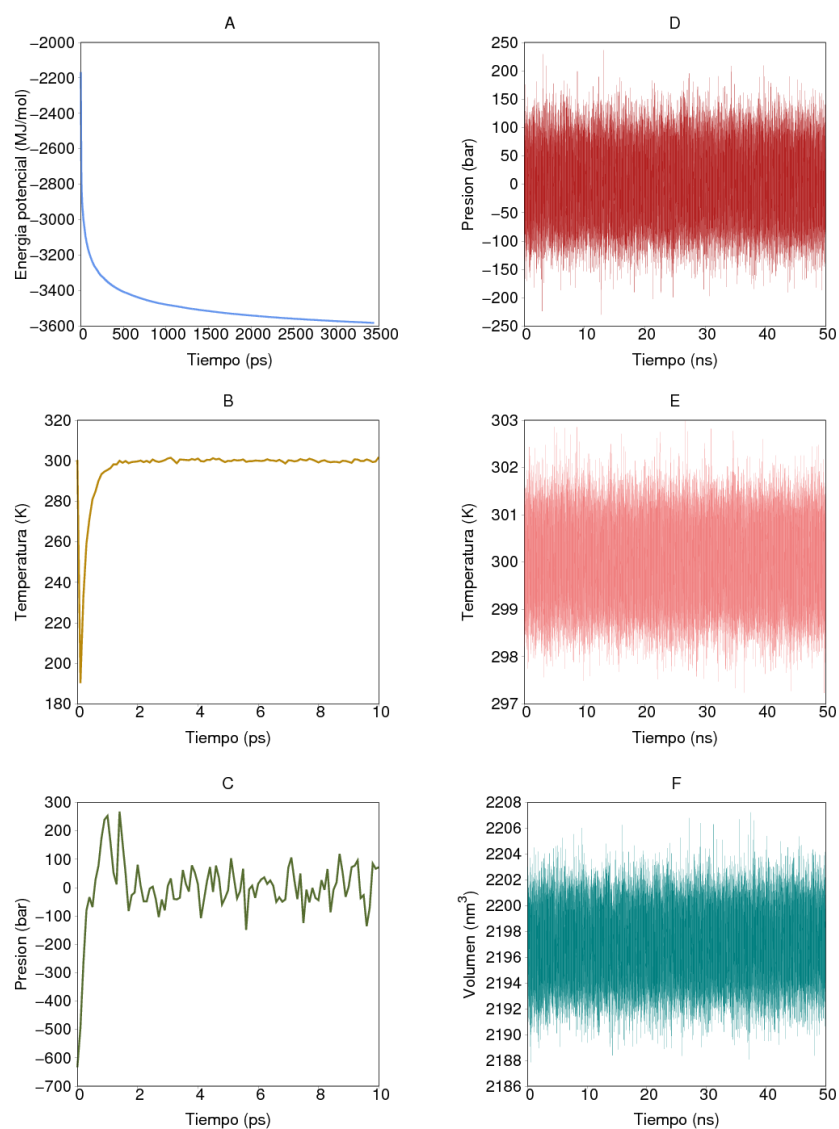


Figura 12.12. Variables de estado a través del tiempo. (A) Energía potencial (MJ/mol) en EM, (B) temperatura (K) en NVT, (C) presión (bar) en NPT, (D) presión (bar) en MD, (E) temperatura (K) en MD, (F) volumen (nm^3) en MD.

XII.6.2. Equilibrio dinámico

El equilibrio dinámico de los sistemas homodiméricos se identificó aplicando DFT a los gráficos de RMSD. Los pasos para seleccionar el intervalo de equilibrio en el control positivo (Figura 12.13) son la aplicación de DFT a la función RMSD, obtener la parte real de la transformada discreta, aplicar la DFT^{-1} y graficar la función resultante. El equilibrio dinámico se consideró cuando la función periódica resultante tiene una amplitud aproximadamente equidistante a un eje horizontal.

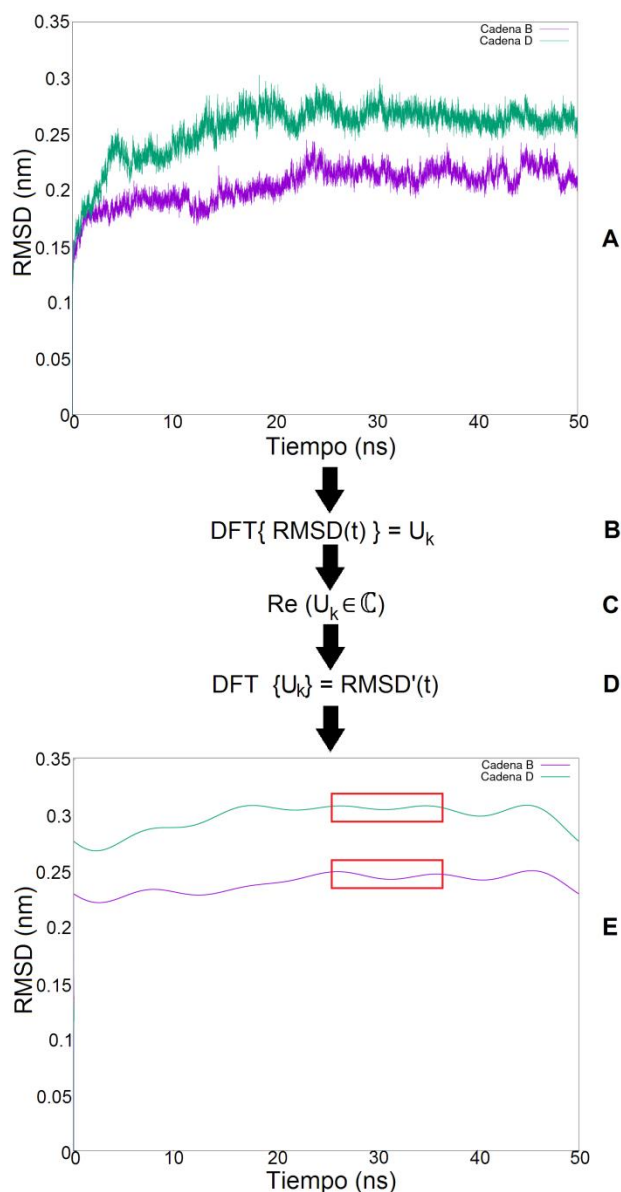


Figura 12.13. Identificación de un equilibrio dinámico en el control positivo. (A) Gráfico de RMSD de la cadena B. (B) Aplicación de DFT. (C) Obtención de la parte real de la transformada. (D) Aplicación de DFT^{-1} . (E) Gráfica de la función suavizada. Se consideró el equilibrio dinámico en el intervalo de tiempo correspondiente al segmento en el rectángulo color rojo.

XII.6.3. Matriz de distancias mínimas

En los sistemas homodiméricos hay cuatro regiones principales de las subunidades que están en contacto cercano: los segmentos 212 a 217, 276 a 297 (que incluye el asa M) y 324 a 332 en la cadena B; y el segmento 111 a 127 en la cadena D, aproximadamente. En la matriz de distancias mínimas para el control positivo (Figura 12.14) se señalan con un rectángulo las regiones de la interfase que forman puentes de hidrógeno.

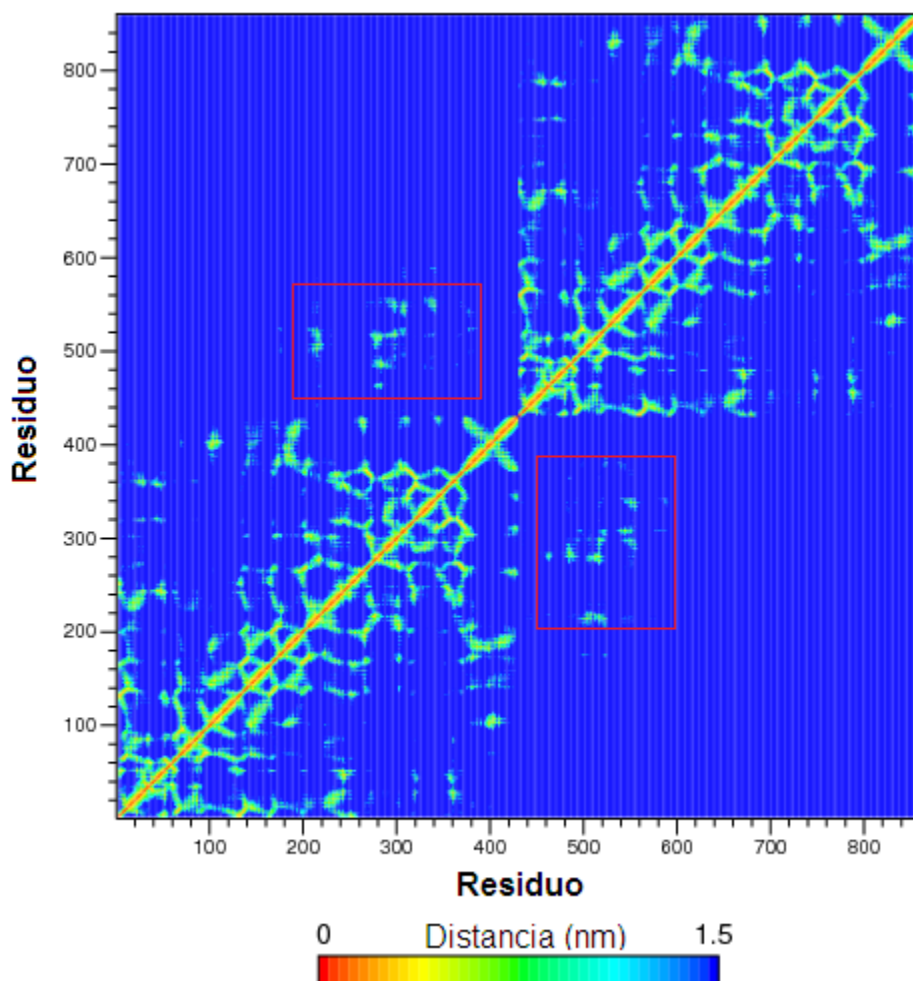


Figura 12.14. Matriz de distancias mínimas entre residuos del dímero β_2 .

XII.7. Energías de solvatación y energías de unión de los ligandos.

Las contribuciones medias de los aminoácidos a la energía de unión por el método MM/PBSA se calcularon para la cadena B y la cadena D. Para el control positivo (Figura 12.15).

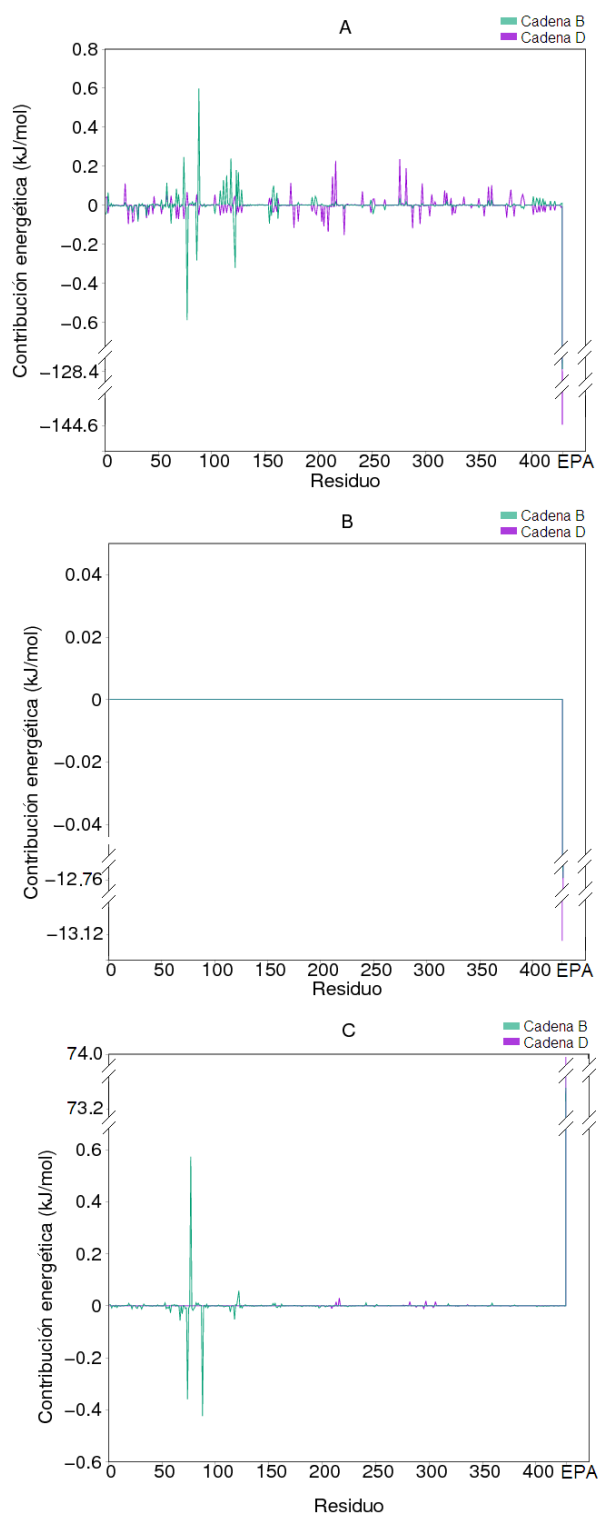


Figura 12.15. Contribuciones de los residuos de aminoácidos a la energía de unión de EPA en el control positivo, por el método MM/PBSA. (A) Contribuciones de mecánica molecular (MM), (B) Contribuciones apolares, (C) contribuciones polares.

XIII. REFERENCIAS

- (1) Kumar, A.; Singh, A.; Ekavali. *Pharmacol. Reports* **2015**, 67 (2), 195–203.
- (2) Tellechea, P.; Pujol, N.; Esteve-Belloch, P.; Echeveste, B.; García-Eulate, M. R.; Arbizu, J.; Riverol, M. *Neurologia* **2015**, 799, 1–10.
- (3) O'Bryant, S. E.; Mielke, M. M.; Rissman, R. A.; Lista, S.; Vanderstichele, H.; Zetterberg, H.; Lewczuk, P.; Posner, H.; Hall, J.; Johnson, L.; Fong, Y.-L.; Luthman, J.; Jeromin, A.; Batrla-Utermann, R.; Villarreal, A.; Britton, G.; Snyder, P. J.; Henriksen, K.; Grammas, P.; Gupta, V.; Martins, R.; Hampel, H. *Alzheimer's Dement.* **2016**, 13 (1), 1–14.
- (4) Snyder, D. S. National Institutes of Health <https://report.nih.gov/NIHfactsheets/>.
- (5) Arendt, T.; Stieler, J. T.; Holzer, M. *Brain Res. Bull.* **2016**, 126, 238–292.
- (6) Carroll, C. M.; Li, Y. M. *Brain Res. Bull.* **2016**, 126, 199–206.
- (7) Buckingham, S.; Jones, A. *Pharmacol. Rev.* **2009**, 61 (1), 39–61.
- (8) Thomas, S. J.; Snowden, J. A.; Zeidler, M. P.; Danson, S. J. *Br. J. Cancer* **2015**, 113 (3), 365–371.
- (9) Drewes, G.; Trinczek, B.; Illenberger, S.; Biernat, J.; Schmitt-Ulms, G.; Meyer, H. E.; Mandelkow, E. M.; Mandelkow, E. *Journal of Biological Chemistry*. 1995, pp 7679–7688.
- (10) Shamitko-Klingensmith, N.; Boyd, J. W.; Legleiter, J. *AIMS Biophys.* **2016**, 3 (2), 261–285.
- (11) Picone, P.; Nuzzo, D.; Caruana, L.; Scafidi, V.; Carlo, M. Di. *Oxid. Med. Cell. Longev.* **2014**, 2014, 1–11.
- (12) A-beta Signalling Pathways www.rndsystems.com.
- (13) Downing, K. H. *Annu. Rev. Cell Dev. Biol.* **2000**, 89–111.
- (14) Leandro-García, L. J.; Leskelä, S.; Landa, I.; Montero-Conde, C.; López-Jiménez, E.; Letón, R.; Cascón, A.; Robledo, M.; Rodríguez-Antona, C. *Cytoskeleton* **2010**, 67 (4), 214–223.
- (15) Kollman, J. M.; Merdes, A.; Mourey, L.; Agard, D. A. *Nat. Rev. Mol. Cell Biol.* **2011**, 12 (11), 709–721.
- (16) Ng, Y. P.; Or, T. C. T.; Ip, N. Y. *Neurochem. Int.* **2015**, 89, 260–270.
- (17) Lee, T. H.; Park, S.; You, M. H.; Lim, J. H.; Min, S. H.; Kim, B. M. *J. Neurochem.* **2016**, 136 (6), 1232–1245.
- (18) Yan, X.; Hu, G.; Yan, W.; Chen, T.; Yang, F.; Zhang, X.; Zhao, G.; Liu, J. *Life Sci.* **2017**, 168, 16–23.
- (19) Zhang, M.; Wang, Y.; Qian, F.; Li, P.; Xu, X. *Biochem. Biophys. Res. Commun.* **2016**, 1, 1–6.
- (20) Field, J. J.; Díaz, J. F.; Miller, J. H. *Chem. Biol.* **2013**, 20 (3), 301–315.
- (21) Rohena, C. C.; Mooberry, S. L. *Nat. Prod. Rep.* **2014**, 31 (3), 335.
- (22) Jordan, M. A.; Wilson, L. *Nat. Rev. Cancer* **2004**, 4 (4), 253–265.

- (23) Brunden, K. R.; Gardner, N. M.; James, M. J.; Yao, Y.; Trojanowski, J. Q.; Lee, V. M. Y.; Paterson, I.; Ballatore, C.; Smith, A. B. *ACS Med. Chem. Lett.* **2013**, *4* (9), 886–889.
- (24) Cartelli, D.; Casagrande, F.; Busceti, C. L.; Bucci, D.; Molinaro, G.; Traficante, A.; Passarella, D.; Giavini, E.; Pezzoli, G.; Battaglia, G.; Cappelletti, G. *Sci. Rep.* **2013**, *3*, 1837.
- (25) Gajewski, M. M.; Tuszynski, J. A.; Barakat, K.; Huzil, J. T.; Klobukowski, M. *Can. J. Chem.* **2013**, *91* (7), 511–517.
- (26) Penazzi, L.; Tackenberg, C.; Ghori, A.; Golovyashkina, N.; Niewidok, B.; Selle, K.; Ballatore, C.; Smith, A. B.; Bakota, L.; Brandt, R. *Neuropharmacology* **2016**, *105*, 84–95.
- (27) Cano-González, L. Síntesis y evaluación in silico de nuevos carbamatos bencimidazólicos como inhibidores de la polimerización de la β -tubulina, 2015.
- (28) Saha, A.; Mohapatra, S.; Kurkute, P.; Jana, B.; Mondal, P.; Bhunia, D.; Ghosh, S.; Ghosh, S. *Chem. Commun. (Camb)*. **2015**, *51* (12), 2249–2252.
- (29) National Center for Biotechnology Information. PubChem Compound Database <https://pubchem.ncbi.nlm.nih.gov>.
- (30) Bickel, U. *NeuroRx* **2005**, *2* (1), 15–26.
- (31) Lagorce, D.; Sperandio, O.; Baell, J. B.; Miteva, M. A.; Villoutreix, B. O. *Nucleic Acids Res.* **2015**, *43* (W1), W200–W207.
- (32) Clark, D. E. *Abstr. Pap. Am. Chem. Soc.* **1999**, *217* (8), U696–U696.
- (33) Abraham, M. H.; Ibrahim, A.; Zhao, Y.; Acree, W. *J. Pharm. Sci.* **2006**, *95* (10), 2091–2100.
- (34) R-Development-Core-Team. R Foundation for Statistical Computing: Vienna, Austria 2008.
- (35) De Oliveira, D. B.; Gaudio, A. C. *Quant. Struct. Relationships* **2001**, *19* (6), 599–601.
- (36) Tropsha, A.; Golbraikh, A. In *Handbook of Chemoinformatics Algorithms*; Faulon, J.-L., Bender, A., Eds.; Chapman & Hall Book, 2010; pp 214–216.
- (37) Wasmuth, E. V.; Lima, C. D. *Nucleic Acids Res.* **2016**, *45* (November 2016), 1–12.
- (38) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (39) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; Thompson, J. D.; Higgins, D. G.; McWilliam, H.; Remmert, M.; So, J. *Mol. Syst. Biol.* **2011**, *7* (539), 1–6.
- (40) Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T. G.; Bertoni, M.; Bordoli, L.; Schwede, T. *Nucleic Acids Res.* **2014**, *42* (W1), 252–258.
- (41) Laskowski, R. A. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
- (42) Benkert, P.; Tosatto, S. C. E.; Schomburg, D. *Proteins-Structure Funct. Bioinforma.* **2008**, *71* (1), 261–277.
- (43) Schrödinger.
- (44) Schrödinger: New York 2017.

- (45) Zoete, V.; Cuendet, M. A.; Grosdidier, A.; Michielin, O. *J. Comput. Chem.* **2010**, *31* (16), 2967–2970.
- (46) Brooks, B. R.; Iii, C. L. B.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M. *J. Comput. Chem.* **2009**, *30* (10), 1545–1614.
- (47) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91* (1–3), 43–56.
- (48) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. *Proc. Python Sci. Conf.* **2016**, No. Scipy, 98–105.
- (49) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2009**, *31* (2), 455–461.
- (50) Kumari, R.; Kumar, R.; Source, O.; Discovery, D.; Lynn, A. *J. Chem. Inf. Model.* **2014**, *54*, 1951–1962.
- (51) Sunseri, J.; Koes, D. R. *Nucleic Acids Res.* **2016**, *44*, 442–448.
- (52) Adams, S. E. *Molecular Similarity and Xenobiotic Metabolism*, University of Cambridge, 2010.
- (53) Ramachandran, N.; Sasisekharan, V. *Adv. Protein Chem.* **1968**, *23*, 283–437.
- (54) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. *J. Comput. Chem.* **2011**, *32*, 2319–2327.
- (55) Touw, W. G.; Baakman, C.; Black, J.; Beek, T. A. H.; Krieger, E.; Joosten, P.; Vriend, G. *Nucleic Acids Res.* **2015**, *43*, 364–368.
- (56) Kabsch, W.; Sander, C. *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features*; 1983; Vol. 22.
- (57) Chakraborty, A.; de Wit, N. M.; van der Flier, W. M.; de Vries, H. E. *Vascul. Pharmacol.* **2016**, *89*, 12–18.
- (58) Rees, D. C.; Johnson, E.; Lewinson, O. *Nat. Rev.* **2009**, *10*, 218–227.
- (59) Friedman, M. H. *Princ. Model. Biol. Transp.* **2008**.
- (60) Mannhold, R.; Poda, G. I.; Ostermann, C.; Tetko, I. V. *J. Pharm. Sci.* **2008**, *98* (3), 861–893.
- (61) Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R.; Lai, L. *J. Chem. Inf. Model.* **2007**, *47* (6), 2140–2148.
- (62) Montgomery C.D., P. E. A. G. G. . **2002**, 61, 65, 79, 80, 82, 83.
- (63) Montgomery, D. C.; Peck, E. A.; Vining, G. G. In *Introducción al análisis de regresión lineal*; México, 2002; pp 117–122.
- (64) Jayakumar, G. S. *Hacettepe Univ. Bull. Nat. Sci. Eng. Ser. B Math. Stat.* **2015**, *44* (1), 165–178.
- (65) Roy, K.; Kar, S.; Das, R. N. In *A Primer on QSAR/QSPR Modeling*; 2015; pp 37–59.
- (66) Tropsha, A.; Gramatica, P.; Gombar, V. K. *QSAR Comb. Sci.* **2003**, No. 3, 69–77.
- (67) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.

- C.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz, V. E.; Cramer, R. D.; Benigni, R.; Yang, C.; Rathman, J. F.; Terfloth, L.; Gasteiger, J.; Richard, A. M.; Tropsha, A. *J. Med. Chem.* **2014**.
- (68) Benkert, P.; Biasini, M.; Schwede, T. *Struct. Bioinforma.* **2010**, 27 (3), 343–350.
- (69) Rangwala, H.; Karypis, G. In *Introduction to Protein Structure Prediction: Methods and Algorithms*; John Wiley & Sons, Inc., 2011; p 301.
- (70) Eisenhaber, F.; Lijnzaad, P.; Argos, P.; Sander, C.; Scharf, M. *J. Comput. Chem.* **1995**, 16, 273–284.
- (71) Baldi, P.; Nasr, R. *J Chem Inf Model* **2011**, 50 (7), 1205–1222.
- (72) Simonson, T. In *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Springer Science & Business Media, 2007.
- (73) Thomas, G. *Medicinal Chemistry. An Introduction*; Wiley & Sons, 2000.
- (74) McMurry, J. In *Química Inorgánica*; International Thomson Editores S.A. de C.V.: México, D.F., 2001; pp 206–210.
- (75) Xu, K.; Schwarz, P. M.; Ludueña, R. F. *Drug Dev. Res.* **2002**, 96 (August 2001), 91–96.
- (76) Dearden, J. C.; T.D.M., C. *Quantitative Structure-Activity Relationships (QSAR) in Drug Design*, 4th ed.; Smith, H. J., Ed.; Taylor & Francis, 2006.
- (77) Rabal-Gracia, O. Tesis doctoral: Herramientas de cribado virtual aplicadas a inhibidores de tirosina quinasas. Contribución al desarrollo del programa PRALINS para el diseño de quimiotecas combinatorias.
- (78) Ertl, P.; Rohde, B.; Selzer, P. *J. Med. Chem.* **2000**, 43 (20), 3714–3717.
- (79) Montgomery, D. C.; Peck, E. A.; Vining, G. G. In *Introducción al análisis de regresión lineal*; Compañía Editorial Continental, 2002; p 1,4.
- (80) Montgomery, D. C.; Peck, E. A.; Vining, G. G. In *Introducción al análisis de regresión lineal*; Compañía Editorial Continental, 2002; p 13,14.
- (81) Montgomery C.D.; Peck, E. A.; Vining, G. G. In *Introducción al análisis de regresión lineal*; Compañía Editorial Continental, 2002; p 61, 62, 65, 74, 78–80, 82,83.
- (82) Freund, J. E. In *Estadística matemática con aplicaciones*; México, 2000; pp 384–386.
- (83) Wasserstein, R. L.; Lazar, N. A.; Wasserstein, R. L.; Lazar, N. A.; Asa, T. *Am. Stat.* **2016**, 70 (2), 129–133.
- (84) Sen, A.; Srivastava, M. In *Regression Analysis: Theory, Methods and Applications*; Casella, G., Fienberg, S., Olkin, I., Eds.; Springer: New York, 1997; p 105.
- (85) Royston, P. *Stat. Comput.* **1991**, 2, 117–119.
- (86) Weisberg, S.; Cook, R. D. *Applied Linear Regression*, Second Edi.; Wiley: New York, 1985.
- (87) Halgren, T. a. *J. Comput. Chem.* **1996**, 17 (5–6), 490–519.
- (88) Kitchen, D.; Decornez, H.; Furr, J.; Bajorath, J. *Nat Rev Drug Discov* **2004**, 3 (11), 935–949.

- (89) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19* (14), 1639–1662.
- (90) de Magalhães, C. S.; Barbosa, H. J. C.; Dardenne, L. E. *Genet. Mol. Biol.* **2004**, *27* (4), 605–610.
- (91) Koenig, A. *CS 447 - Adv. Top. Artif. Intell.* **2002**, 1–8.
- (92) Scott, W. R. P.; Hu, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kru, P.; Gunsteren, W. F. Van. *J. Phys. Chem.* **1999**, *103*, 3596–3607.
- (93) Abraham, M.; Hess, B.; van der Spoel, D.; Lindahl, E. The GROMACS development teams at the Royal Institute of Technology and Uppsala University: Uppsala 2014.
- (94) Jordaan, J. A.; Živanović, R. *Measurement* **2006**, *39*, 451–457.
- (95) Maxwell, J. C. *Philosophical Transactions of the Royal Society of London*. Londres 1867, pp 49–88.
- (96) Jazar, R. N. In *Advanced Dynamics, Rigid Body, Multibody, and Aerospace Applications*; John Wiley & Sons, Inc.: Hoboken, 2011; pp 887–932.
- (97) Haaser, N. B.; LaSalle, J. P.; Sullivan, J. A. In *Análisis matemático 1*; Editorial Trillas, S.A.I: México, D.F., 1970.
- (98) Landau, L.; Lifshitz, E. In *Curso Abreviado de Física Teórica: Mecánica y Electrodinámica*; MIR: Moscú, 1971; p 57,58.
- (99) Darden, T.; York, D.; Pedersen, L.; Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98* (10089), 1–5.
- (100) Essmann, U.; Perera, L.; Darden, M. L. B.; Lee, H.; Pedersen, L. G.; Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103* (8577), 8577–8593.
- (101) Kolmogorov, A. N.; Fomin, S. V. In *Elementos de la teoría de funciones y del análisis funcional*; MIR: Moscú, 1972; p 130,131.
- (102) Aleksandrovich, Y.; Syromyatnikov, V. N. In *Phase Transitions and Crystal Symmetry*; Springer Science & Business Media, 2012; p 12.
- (103) Reifenberg, R. Inter-Molecular Forces:Keesom Force
<https://www.physics.purdue.edu/~reifebr>.
- (104) Gross, J.; Engineering, C. P.; Ag, B.; Gic, P. Q. *Am. Inst. Chem. Eng. J.* **2005**, *51* (9), 2556–2568.
- (105) van Oss, C. J. *J. Mol. Recognit.* **2003**, *16*, 177–190.
- (106) Kolmogorov, A. N.; Fomin, S. V. In *Elementos de la teoría de funciones y del análisis funcional*; MIR: Moscú, 1972; pp 464–468.
- (107) O’Neil, P. V. In *Matemáticas Avanzadas para Ingeniería*; CENGAGE Learning: México, 2015; pp 49, 52, 99.
- (108) O’Neil, P. V. In *Matemáticas Avanzadas para Ingeniería*; CENGAGE Learning: México, 2015; p

108, 110, 131,142.

- (109) Schrödinger, L. Schrödinger: Scientific leader in life sciences and materials research <https://www.schrodinger.com/>.
- (110) Huheey, J. E.; Keiter, E. A.; Keiter, R. L. In *Química Inorgánica*; Oxford University Press: México, 1997.
- (111) Freund, J. E.; Miller, I.; Miller, M. In *Estadística matemática con aplicaciones*; México, 2000; pp 539–542.
- (112) Åqvist, J.; Medina, C.; Smauelsson, J.-E. *Protein Eng.* **1994**, 7 (3), 385–391.