

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMÁTICAS



Máster en Ingeniería Matemática
CURSO 2017/2018

ESTIRANDO PROTEÍNAS

Diego Botija Slavicek

Dirección
Ana Carpio

ÍNDICE

I. Introducción	3
II. Modelo teórico	6
1. Derivación del modelo	6
2. Adimensionalización	8
III. Resolución numérica	9
IV. Resultados numéricos para la dinámica determinista	12
1. Puntos de equilibrio	12
2. Experimentos de fuerza-extensión	13
3. Experimentos de fuerza-abrazadera	15
4. Estimación del tiempo	18
V. Resultados numéricos para la dinámica estocástica	20
VI. Conclusiones	23
VII. Códigos Matlab	24
1. Resolución numérica de la curva fuerza-extensión	24
2. Resolución numérica de la dinámica determinista	26
3. Resolución numérica de la dinámica estocástica	28
Bibliografía	30

I. INTRODUCCIÓN

Las proteínas realizan una gran cantidad de funciones en los organismos. Para comprender el funcionamiento estructural y contráctil de las proteínas es necesario estudiar sus propiedades elásticas. La elasticidad de los tejidos resultan de los procesos de plegamiento y extensión de ciertas proteínas unidas por sus extremos a estructuras rígidas. El estudio de proteínas aisladas aumenta la comprensión de las propiedades y funciones de los tejidos, como la contracción muscular. Las proteínas modulares, o poliproteínas, como la titina (que interviene en la contracción muscular) y la ubiquitina están formadas por varios dominios proteicos, zonas de la proteína que pueden funcionar con independencia de las demás, unidos por enlaces peptídicos. Estos módulos de las proteínas pueden extenderse y plegarse de forma casi independiente.

El desarrollo de técnicas experimentales ha permitido el estudio de las propiedades elásticas de proteínas aisladas, fuera del medio de los tejidos. En la microscopía de fuerza atómica (AFM) uno de los extremos de la proteína se une a una estructura fija y el otro a la punta de un microscopio. La punta de este microscopio es móvil y es capaz de medir fuerzas del orden de piconewton (pN). Cuando se mueve la punta para estirar la proteína esta se extiende cuando está sometida a una fuerza mayor a un cierto umbral superior. Cuando el microscopio disminuye la fuerza aplicada por debajo de un umbral inferior los módulos vuelve a plegarse. En el proceso de extensión y plegamiento interviene el ruido térmico, debido a la temperatura ambiente del sistema, por lo que tienen una naturaleza estocástica. Estos procesos son completamente reversibles.

Con este montaje experimental aplicado a una proteína modular se pueden realizar experimentos donde se controla el movimiento de la punta y se mide la fuerza aplicada en función de la longitud de la proteína para obtener curvas de fuerza-extensión (FEC). Estas curvas han sido estudiadas y su comportamiento es bien conocido: la curva tiene forma de dientes de sierra, donde los saltos se producen cuando uno de los módulos cambia su configuración de plegado a extendido. Otro experimento

que se puede realizar es el de fuerza-abrazadera (*force-clamp*) en el que se controla la fuerza que se aplica. Típicamente se parte de una fuerza inicial pequeña o nula que se aumenta rápidamente hasta una fuerza grande que se mantiene hasta que la proteína se extiende en su totalidad. Posteriormente se disminuye la fuerza abruptamente a un valor fijo menor, para estudiar la extensión y plegamiento de la proteína durante todo el proceso.

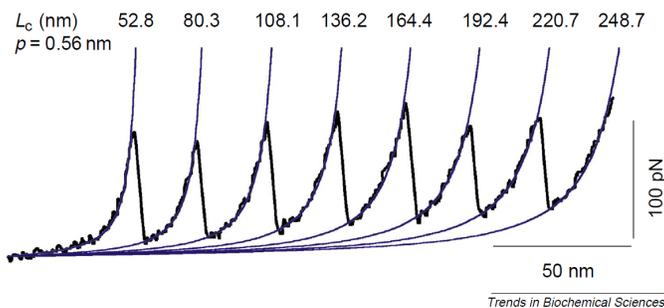


Figura 1: Curva de fuerza-extensión de la proteína modular Tenascina C obtenida en [4], con un ajuste con el modelo de la varilla elástica con longitud de contorno L_c indicada para cada pico y longitud de persistencia $P = 0,56 \text{ nm}$.

En la figura 1 de [4] se puede ver la curva fuerza-extensión de un experimento con el montaje experimental explicado, para la proteína modular Tenascina C, donde se observan claramente cuando se produce un salto debido a la extensión de uno de los módulos.

En la figura 2 de [2] se observa el resultado de un experimento de *fuerza-abrazadera* para un solo módulo de una proteína. En la figura se puede notar que cuando se aplica una fuerza grande el módulo se extiende un poco y se mantiene en ese estado durante un tiempo hasta que se extiende completamente. Una vez tiene la extensión máxima y se disminuye la fuerza aplicada, la longitud disminuye en d_1 y se mantiene en un estado metaestable. Después de un tiempo que parece aleatorio el módulo disminuye en longitud d_2 y llega a un estado plegado estable.

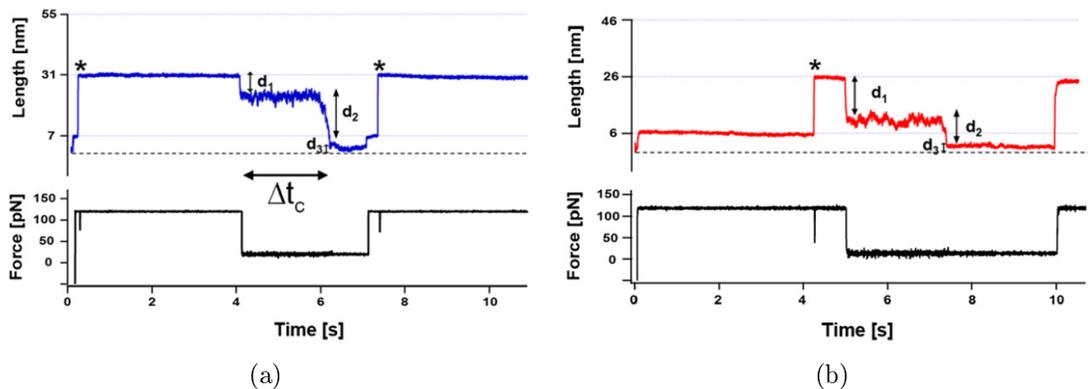


Figura 2: Longitud de (a) un monómero de la proteína I27 y (b) un monómero de la proteína ubiquitina a lo largo del tiempo y la fuerza en cada instante a la que están sometidos, obtenidos en [2].

Para realizar un estudio matemático de este fenómeno hace falta un modelo teórico con el que se obtengan resultados que se puedan comparar a los obtenidos en los experimentos. Con el modelo que se plantea a continuación se obtiene un potencial para los módulos, que están unidos entre sí por un enlace con un potencial que se puede aproximar por uno parabólico. El comportamiento de los módulos se puede simular aplicando la ecuación de Langevin, con una parte determinística y otra estocástica, que se puede integrar para obtener la dinámica de la proteína para unas condiciones dadas similares a las de los experimentos [2].

El objetivo del trabajo es revisar los modelos de [1] y aprender a implementarlos y simularlos, analizando por último los resultados obtenidos. Se han implementado los modelos en MATLAB para hacer simulaciones de los experimentos mencionados, haciendo una interpretación de la extensión y plegamiento a través de frentes de onda. Se han comparado los resultados con otros obtenidos experimentalmente [2,4,5] y teóricamente [1].

II. MODELO TEÓRICO

1. Derivación del modelo

El tiempo característico en el que ocurren los procesos de extensión y plegamiento es mucho mayor que el tiempo de relajación de la tensión dentro de los módulos, por lo tanto se puede suponer que cada módulo está en equilibrio mecánico en cada instante. Para estudiar el comportamiento elástico de los módulos hay que examinar su energía libre, que es la parte de la energía que puede intercambiar con el entorno en forma de trabajo. La energía libre presente en procesos elásticos de una proteína tiene dos componentes: un término entrópico que tiene en cuenta la elasticidad de la cadena y un término entálpico de interacción a corta distancia entre los aminoácidos.

La extensión de una proteína está muy bien descrita por el modelo de la varilla elástica (WLC). Este modelo utilizado en física de polímeros describe una fibra continuamente flexible. Una fórmula de interpolación que describe la extensión u de un polímero [3] con longitud de contorno L_c y longitud de persistencia P sometido a una fuerza F por sus extremos es

$$\frac{FP}{k_B T} = \frac{1}{4} \left(1 - \frac{u}{L_c}\right)^{-2} - \frac{1}{4} + \frac{u}{L_c}, \quad (1)$$

donde k_B es la constante de Boltzmann y T la temperatura absoluta. De esta expresión podemos sacar un potencial, fijando $V_{WLC}(u=0) = 0$,

$$V_{WLC} = \int_0^u -F(\hat{u}) d\hat{u} = \frac{k_B T L_c}{4P} \left(\frac{1}{1 - \frac{u}{L_c}} - 1 - \frac{u}{L_c} + \frac{2u^2}{L_c^2} \right). \quad (2)$$

Este va a ser el potencial que describa la parte entrópica del proceso.

El término entálpico se puede aproximar por un potencial de Morse

$$V_M = U_0 \left[\left(1 - e^{-2b(u-R_c)/R_c}\right)^2 - 1 \right], \quad (3)$$

donde U_0 , R_c y b son los parámetros a ajustar que definen la profundidad, la posición

y la dispersión del pozo de potencial. Este término determina el mínimo entálpico de la proteína plegada, que crece muy rápido si la proteína intenta plegarse más pero permite que se expanda si tiene la energía suficiente.

El potencial total es

$$V(u) = V_M + V_{WLC} = U_0 \left[\left(1 - e^{-2b(u-R_c)/R_c}\right)^2 - 1 \right] + \frac{k_B T L_c}{4P} \left(\frac{1}{1 - \frac{u}{L_c}} - 1 - \frac{u}{L_c} + \frac{2u^2}{L_c^2} \right). \quad (4)$$

Para la ubiquitina se toman los parámetros que toman en [1] $P = 0,28 \text{ nm}$, $L_c = 30 \text{ nm}$, $U_0 = 200 \text{ pN} \cdot \text{nm}$ ($\sim 48k_B T$), $R_c = 4 \text{ nm}$, $b = 2$, $T = 300 \text{ K}$. Como parte de la estructura terciaria de la proteína modular los módulos están unidos por un potencial que se aproxima al de un oscilador armónico $V_h = \frac{k}{2}(u_j - u_{j-1})^2$, como se sugiere en [1], para dos módulos sucesivos $j - 1$ y j , con una constante elástica $k = 0,065 [F]/L_c$.

Cada módulo satisface la ecuación de Langevin sobreamortiguada [2]

$$\gamma \dot{u}_j = -\frac{\partial}{\partial u_j} \mathcal{G}(\mathbf{u}, F, T) + \sqrt{2D} \gamma \xi_j(t), \quad (5)$$

donde $\mathcal{G}(\mathbf{u}, F, T) = \sum_{j=1}^N [V(u_j) - F u_j + \frac{k}{2}(u_j - u_{j-1})^2]$ es el potencial total para una proteína de N módulos, $D = k_B T / \gamma$ es el coeficiente de difusión y ξ_j es ruido blanco. La ecuación de Langevin es una ecuación diferencial estocástica que describe la evolución de partículas en un potencial, con un ruido producido por movimiento browniano. La variable que aparece en la ecuación es una variable colectiva que varía más lentamente que las variables microscópicas de las que depende. En este caso la variable colectiva es la longitud de cada módulo proteico que depende del movimiento estocástico de cada pequeña sección de la cadena debido a la energía térmica. La longitud total de la proteína modular es $L = \sum_{j=1}^N u_j$. Se asumen uniones infinitamente rígidas en la estructura fija y la punta con la proteína, de forma que $u_0 = u_1$ y $u_{N+1} = u_N$.

2. Adimensionalización

Para trabajar en la escala en la que se desarrollan los experimentos y no preocuparnos por las unidades físicas se pueden adimensionalizar las ecuaciones. Se puede comprobar que las variables $u^* = \frac{u}{L_c}$ y $t^* = \frac{[F]}{\gamma L_c} t$, con $[F] = 100pN$, son adimensionales. Haciendo los cambios de variables, teniendo en cuenta que la variable aleatoria $\xi(t)$ escala con la raíz del tiempo, tenemos

$$\begin{aligned} \gamma L_c \frac{[F]}{\gamma L_c} \frac{du_j^*}{dt^*} &= -\frac{1}{L_c} \frac{\partial}{\partial u_j^*} \sum_{j=1}^N \left[V(u_j^*) - FL_c u_j^* + \frac{k}{2} (u_j - u_{j-1})^2 \right] + \sqrt{\frac{2D[F]}{\gamma L_c}} \gamma \xi_j(t^*) \\ \Rightarrow \frac{du_j^*}{dt^*} &= F^* - \frac{dV^*}{du_j^*} + \kappa (u_{j+1}^* + u_{j-1}^* - 2u_j^*) + \sqrt{2\theta} \xi_j(t^*) , \end{aligned}$$

donde

$$F^* = \frac{F}{[F]}, \quad V^*(u^*) = \mu \left\{ [1 - e^{-\beta(u^* - \rho)}]^2 - 1 + A \left(\frac{1}{1 - u^*} - 1 - u^* + 2(u^*)^2 \right) \right\} ,$$

$\mu = U_0/(L_c[F])$, $\beta = 2bL_c/R_c$, $\rho = R_c/L_c$, $\kappa = kL_c/[F]$, $A = k_B T L_c/(4PU_0)$, $\theta = k_B T/([F]L_c) = 0,0014$. La escala de tiempo, $[t] = \gamma L_c/[F]$, depende de γ que es función de la constante de difusión D . De esta constante van a depender los tiempos característicos de extensión y plegamiento. En el resto del trabajo las variables que aparecen sin unidades se van a referir a las variables adimensionales definidas. La ecuación de Langevin adimensionalizada queda de forma más clara como

$$\dot{u}_j = F - V'(u_j) + \kappa(u_{j+1} + u_{j-1} - 2u_j) + \sqrt{2\theta} \xi_j(t) . \quad (6)$$

III. RESOLUCIÓN NUMÉRICA

Para el estudio de este sistema uno de los primeros pasos es estudiar los estados estables de los módulos, que se corresponden con las extensiones para las cuales el potencial menos la energía elástica proporcionada $V(u) - Fu$ es mínimo. Se puede encontrar los mínimos de esta función de una variable fácilmente con un algoritmo de optimización numérico. Para estudiar el número de mínimos de esta función que depende del valor del parámetro F es interesante trabajar con la derivada de la función $V'(u) - F$. Los puntos de equilibrio del sistema se corresponden con los ceros de esta función, con derivada positiva los que se corresponden con mínimos (equilibrios estables) y derivada negativa los que se corresponden con el máximo (equilibrio inestable), para los casos en los que exista. En el código implementado se calculan los ceros con la función *fzero* de MATLAB, que es un algoritmo de búsqueda de ceros que mezcla el método de bisección, secante e interpolación cuadrática inversa, a partir de un punto inicial.

En la simulación de la curva fuerza-extensión para cada instante t_i , en el que la proteína tiene una longitud L_i determinada, la proteína se encuentra en equilibrio mecánico, en la configuración de mínima energía con la condición que tenga extensión L_i . Para encontrar esta configuración en cada instante será necesario encontrar el mínimo para la suma de las energías libres de los N módulos, que depende de la configuración de extensiones $\vec{u} = (u_1, \dots, u_N)$. La minimización de la función tiene las restricciones de uniones infinitamente rígidas $u_0 - u_1 = 0$, $u_{N+1} - u_N = 0$ y la restricción de longitud total determinada $\sum_{i=1}^N u_i = L$. Se ha probado la función *fmincon* probando diferentes algoritmos entre los que permite elegir, pero mostraban inestabilidad en la solución, pasando a veces a valores numéricos muy grandes. Finalmente se han implementado las restricciones como términos adicionales en la función, de la forma $c_1(u_0 - u_1)^2 + c_2(u_{N+1} - u_N)^2 + c_3(\sum_{i=1}^N u_i - L)^2$ con coeficientes c_1, c_2, c_3 relativamente grandes, de forma que al minimizar la nueva función se fuerzan las restricciones y la solución no cambia respecto a la original cuando se cumplen. Para buscar numéricamente el mínimo de esta función de N variables se

ha utilizado la función *fminsearch* de MATLAB, que aplica el método de Nelder-Mead para la búsqueda: un algoritmo que usa un simplex de $N + 1$ vértices que va moviendo sus vértices dependiendo del valor de la función en ellos. Esta función encuentra un mínimo local cercano al punto inicial de búsqueda proporcionado. La búsqueda del mínimo se va a realizar para cada instante durante un periodo, en el que se va aumentando poco a poco la extensión de la proteína. Como la extensión se hace lentamente en cada búsqueda del mínimo se introduce como punto inicial en el algoritmo la solución del instante anterior, porque el nuevo mínimo será cercano al anterior.

La resolución numérica de la ecuación de Langevin se ha realizado con métodos explícitos, de orden 1 (Euler explícito) y de orden 2 (*ode23t* de MATLAB). El método de Euler es el más sencillo, que para una ecuación diferencial $X'(t) = a(X, t)$ da una solución $X^{n+1} = X^n + a(X^n, t^n)\Delta t$, para $n = 0, \dots, N_t$, con $t^{n+1} = t^n + \Delta t$ y $X^n = X(t^n)$. Aplicado a la ecuación (6) da la solución numérica

$$u_j^{n+1} = u_j^n + [F - V'(u_j^n) + \kappa(u_{j+1}^n + u_{j-1}^n - 2u_j^n)]\Delta t , \quad (7)$$

para $j = 1, \dots, N$, con error $O(\Delta t)$. El método *ode23t* es un método explícito adaptativo de par encajado que utiliza la regla del trapecio. Muchos de los métodos de la familia *ode* no eran estables para este problema o presentaban oscilaciones numéricas en el resultado. Con el paso de tiempo usado con el método Euler explícito $\Delta t = 0,001$ el método era estable, el resultado era visualmente indistinguible del obtenido con *ode23t* y el tiempo de cálculo era corto.

Como se ha visto que en la resolución numérica de la ecuación de Langevin determinista con el método de Euler se obtiene visualmente el mismo resultado que con métodos numéricos de orden superior, se ha resuelto la ecuación estocástica con el método de Euler-Maruyama [6], que es una generalización del método de Euler a ecuaciones diferenciales estocásticas. En el método de Euler-Maruyama, para resolver ecuaciones diferenciales estocásticas, para procesos de Itô, que tienen la forma

$$dX_t = a(X_t)dt + b(X_t)dW_t ,$$

se obtiene la solución numérica Y_t mediante el método iterativo

$$Y_{n+1} = Y_n + a(Y_n)\Delta t + b(Y_n)\Delta W_n, \quad Y_0 = X_0 .$$

Existe un método de orden superior para la resolución numérica de ecuaciones diferenciales estocásticas, el método de Milstein, que para casos donde en la ecuación diferencial b no depende de X_t , como es en este caso, el método es exactamente igual al método de Euler-Maruyama. El método aplicado en este caso tiene entonces la precisión de método de orden superior de $O(\Delta t)$, en vez de la precisión que tiene el método de Euler-Maruyama en general de $O(\Delta t^{1/2})$.

Para la ecuación (6) este método nos da una una solución numérica para una realización de forma

$$u_j^{n+1} = u_j^n + [F - V'(u_j^n) + \kappa(u_{j+1}^n + u_{j-1}^n - 2u_j^n)]\Delta t + \sqrt{2\theta}\Delta\xi_n , \quad (8)$$

donde $\Delta\xi_n \sim \mathcal{N}(0, 1)$.

IV. RESULTADOS NUMÉRICOS PARA LA DINÁMICA DETERMINISTA

1. Puntos de equilibrio

En primer lugar se va a estudiar el sistema sin ruido. En esta sección se va a ver que los procesos de extensión y plegado de las proteínas aparecen en la parte determinística de las ecuaciones.

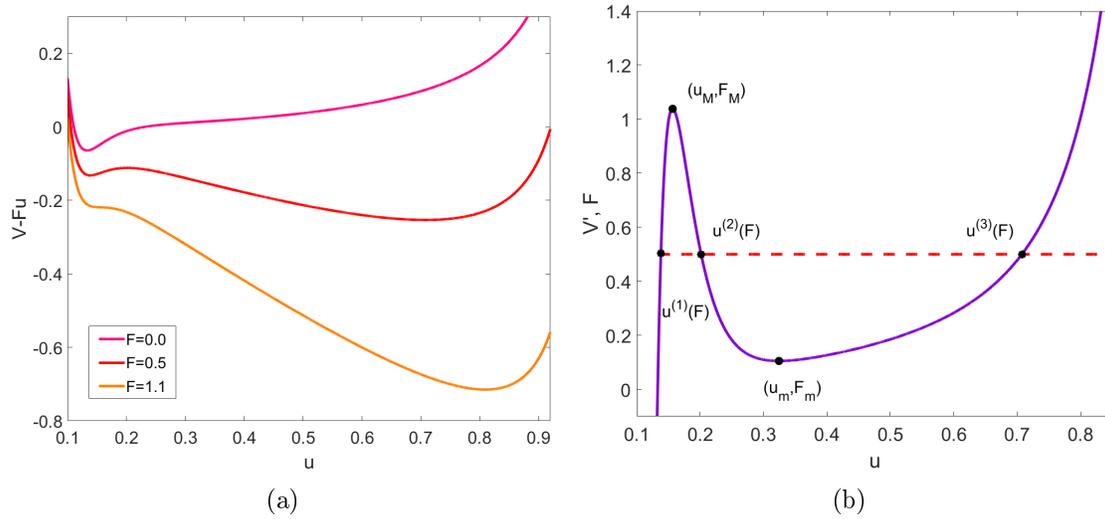


Figura 3: (a) Potencial total $V(u) - Fu$ para fuerzas de $F = 0 \text{ pN}$, 50 pN , 110 pN , de arriba a abajo, con los parámetros para la ubiquitina $L_c = 30 \text{ nm}$, $P = 0,28 \text{ pN} \cdot \text{nm}$, $R_c = 4 \text{ nm}$, $b = 2$, $T = 300 \text{ K}$. (b) Campo de fuerza $V'(u)$ frente a u y fuerza constante $F = 50 \text{ pN}$. Existen tres soluciones a la ecuación $V'(u) = F$, $u^{(1)}(F) < u^{(2)}(F) < u^{(3)}(F)$ para el rango de fuerzas $F_m = 10,5 \text{ pN} < F < F_M = 103,8 \text{ pN}$.

En la figura 3 (a) se puede observar que si no se aplica ninguna fuerza externa, $F = 0 \text{ pN}$, sólo existe un mínimo que se corresponde con el mínimo entálpico en el que el módulo permanece plegado. En el caso en el que se aplica una fuerza de 50 pN hay dos mínimos y en caso de que no hubiera ruido el módulo permanecería en el estado en el que se encontrara, plegado o extendido, para siempre. En el caso en el que se aplica una fuerza de 110 pN solo persiste el mínimo entrópico en el que el módulo se encuentra extendido. En la figura 3 (b) se puede ver que $V'(u)$ tiene una forma cúbica y para qué valores de la fuerza existe un cero, que se corresponde

con una única solución de estabilidad, y para qué valores existen tres ceros, que se corresponden con dos soluciones estables, con extensiones $u^{(1)}(F)$ y $u^{(3)}(F)$, y una inestable en $u^{(2)}(F)$. El perfil de la curva de extensión de un módulo que pasa de un equilibrio a otro se llama frente de onda. Para la región de fuerzas con tres soluciones en el potencial hay una barrera separando los estados plegado y extendido, haciendo que los módulos se mantengan en la configuración en la que se encuentran. Para estas fuerzas tenemos un frente de ondas estacionario, puesto que los módulos de la proteína no pasarían de un estado de equilibrio a otro en un tiempo finito. Cuando se cambia la fuerza a una en la que solo hay un punto de equilibrio, entonces los módulos se mueven hacia él si no se encuentran ya en el equilibrio. En este caso el frente de onda deja de ser estacionario y pasa a moverse.

2. Experimentos de fuerza-extensión

En un experimento de fuerza-extensión de una proteína sometida a una fuerza, con un número determinado de módulos extendidos, se encuentra en equilibrio en la configuración más estable, en la que la energía libre es mínima. Cuando se aumenta la fuerza lentamente el equilibrio se mueve hacia configuraciones de extensión total mayor, hasta que la fuerza en los módulos supera el umbral en el que se pasa a tener una única solución estable, en una configuración extendida. En este momento uno de los módulos se extiende y rápidamente se alcanza un nuevo equilibrio en el que la longitud total disminuye. La curva de fuerza-extensión que se observa en este experimento tiene forma de dientes de sierra, con tantos saltos como módulos tiene la proteína. Para obtener una curva simulada con el modelo teórico se despeja la expresión de la fuerza de la suma de la ecuación de Langevin (6) (sin el ruido) de todos los módulos,

$$F(\mathbf{u}) = \frac{1}{N} \sum_{i=1}^N [\dot{u}_j + V'(u_j) - \kappa(u_{j+1} + u_{j-1} - 2u_j)] . \quad (9)$$

Como los módulos se encuentra en equilibrio mecánico en cada instante $\dot{u}_j = 0$ para todo j . La configuración de \vec{u} en cada instante se obtiene minimizando la energía

libre de la proteína, de la forma en la que se ha explicado en el apartado de resolución numérica.

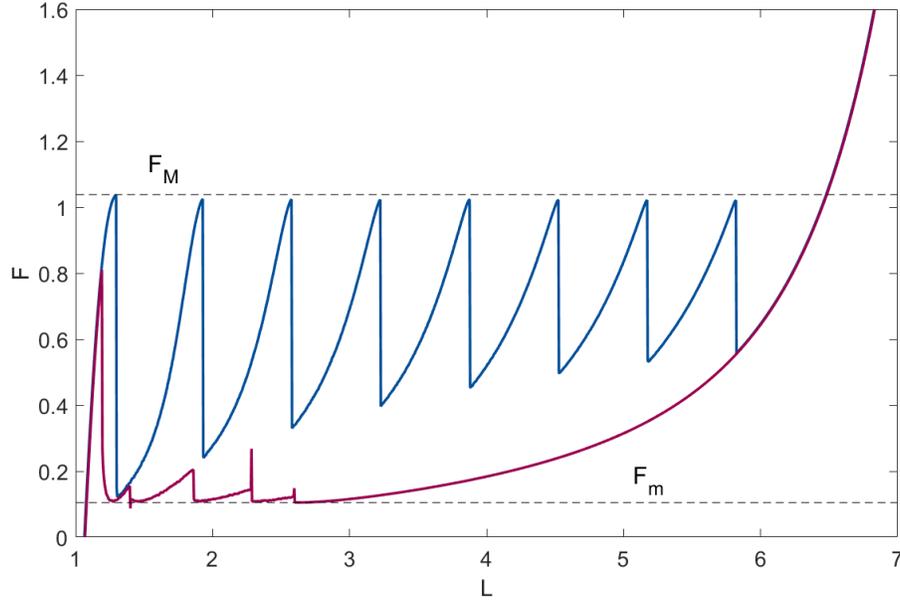


Figura 4: Curva de fuerza-extensión de una proteína de $N = 8$ módulos a partir de la expresión de la fuerza obtenida de suma de las ecuaciones de Langevin de cada módulo (7), para una proteína extendiéndose (azul) de $L = 1[L] = 30 \text{ nm}$ hasta $L = 7[L] = 210 \text{ nm}$ y plegándose (rojo) desde la extensión final a la inicial.

En la figura 4 puede verse que en la curva de extensión, en azul, cuando se alcanza la fuerza $F_M = 1,038[F] = 103,8 \text{ pN}$, a partir la cual el mínimo entrópico es el único que existe, el primer módulo se extiende. Para los siguientes módulos la fuerza crítica a la que se extienden los siguientes es un poco menor que F_M , como veremos más adelante. Esto ocurre tantas veces como módulos tiene la proteína. En la curva de plegamiento, en rojo, todos los módulos permanecen extendidos hasta que la fuerza desciende hasta $F_m = 0,105[F] = 10,5 \text{ pN}$, desde la que unicamente existe el mínimo entrópico, y uno o varios módulos se pliegan. Este resultado es muy parecido a los mostrados en la figura 3, sin la componente del ruido.

3. Experimentos de fuerza-abrazadera

En un experimento de fuerza-abrazadera se parte del mismo montaje experimental con la proteína sometida a una fuerza inicial F_{in} pequeña. Esa fuerza se incrementa bruscamente a una fuerza F_p que se mantiene durante un tiempo determinado y se vuelve modificar bruscamente a una fuerza F_f menor. En función de la fuerza F_p todos los módulos de la proteína se extienden simultáneamente o lo hacen en una sucesión de extensiones escalonadas. De la misma forma, dependiendo de la fuerza F_f la proteína se puede plegar directamente o de forma escalonada. Se puede estudiar el tipo de plegamiento y extensión en función de estas fuerzas.

Inicialmente se tiene la proteína con todos sus módulos plegados, con una extensión $u^{(1)}(F_{in})$. Si la fuerza F_p que se aplica a continuación es mayor que F_M entonces no existe un mínimo en la configuración plegada para ninguno de los módulos. En el instante en el que se incrementa la fuerza a F_p los módulos se extienden hasta el estado $u^{(1)}(F_M)$, durante un breve tiempo. Después de ese tiempo todos los módulos se extienden simultáneamente hasta llegar a la configuración extendida de extensión $u^{(3)}(F_p)$. En el momento en el que la fuerza se reduce a F_f , todos los módulos se encuentran extendidos. Si $F_f < F_m$ la situación es muy similar a la de extensión explicada. Los módulos se encojen hasta la extensión $u^{(3)}(F_m)$, donde pasan un tiempo hasta que todos los módulos se pliegan simultáneamente hasta llegar a una extensión $u^{(1)}(F_f)$. Más adelante veremos que en este caso el tiempo que pasan en el estado $u^{(3)}(F_m)$ es mayor.

En la figura 5 se muestra la simulación sin ruido para una proteína sometida a fuerzas de $F_{in} = 0$, $F_p = 110 \text{ pN}$ y 10 pN consecutivamente, con suficiente tiempo en cada fase para que alcance el equilibrio. Todos los módulos se extienden simultáneamente cuando se aplica F_p en un tiempo breve, y se pliegan a la vez cuando se reduce la fuerza a F_f , en un tiempo mucho más largo. Este resultado se correspondería con el experimento de la figura 2 para una proteína modular. Como todos los módulos se extienden y se pliegan al mismo tiempo el comportamiento es similar al del experimento.

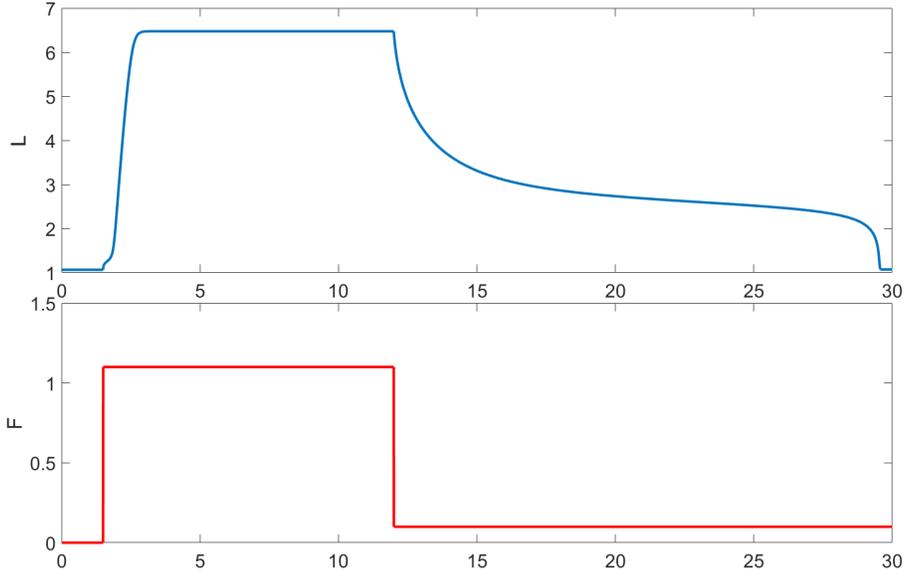


Figura 5: (a) Longitud total de la proteína, inicialmente con extensión $u^{(1)}(F_{in} = 0)$ en todos los módulos, a tiempo t sometida a la fuerza $F(t)$. (b) Fuerza aplicada a tiempo t , que inicialmente es $F_{in} = 0$, tiene un máximo $F_p = 1,1 = 110 \text{ pN}$ mayor que F_M y un valor final $F_f = 0,1 = 10 \text{ pN}$ menor que F_m .

En el caso en el que la fuerza F_p es menor que F_M pero mayor que una cierta fuerza crítica F_{c2} cercana se produce una extensión escalonada. Consideramos una configuración en el que el primer módulo está extendido y los siguientes plegados. El segundo módulo siente una fuerza debida a la unión entre módulos, que habíamos modelado como la interacción de un muelle, con el primer módulo mayor que con el tercero, que tiene su misma longitud. Esta fuerza hace que el segundo módulo se vaya expandiendo y modificando la fuerza umbral necesaria para pasar al estado extendido, a una fuerza F_{c2} . Como la fuerza aplicada $F_p > F_{c2}$ la configuración es inestable y el segundo módulo acaba extendiéndose, hasta una longitud $u^{(3)}(F_p)$. A continuación el segundo módulo se encuentra en la configuración en la que se encontraba el primero inicialmente, y el tercer módulo en el que estaba el segundo. Este suceso va ocurriendo módulo a módulo con un cierto tiempo característico, produciendo un frente de onda que se propaga, hasta que todos los módulos se encuentran en la configuración extendida.

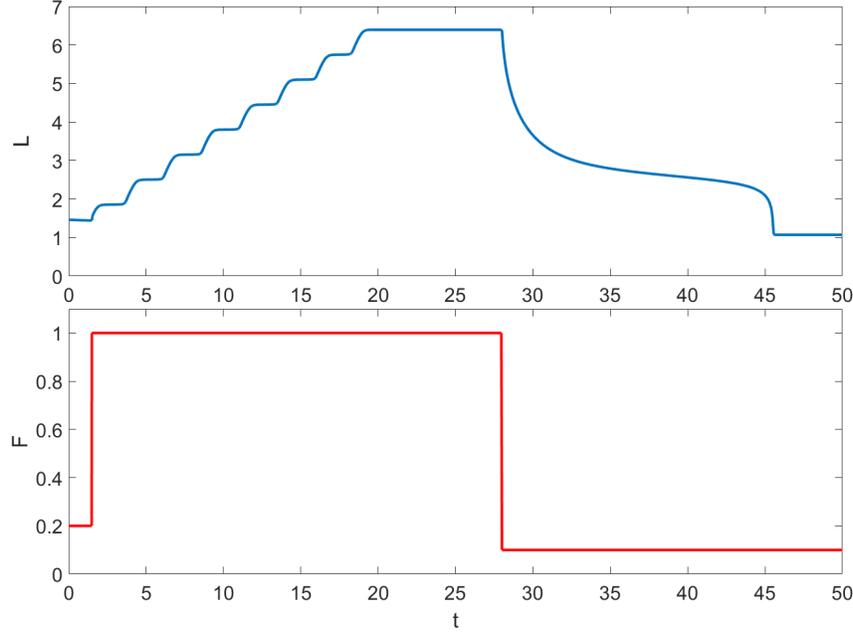


Figura 6: (a) Longitud total de la proteína, inicialmente con extensión $u^{(3)}(F_{in} = 0, 2)$ para $j < 2$ y $u^{(1)}(F_{in} = 0, 2)$ para el resto, a tiempo t sometida a la fuerza $F(t)$. (b) Fuerza aplicada a tiempo t , que inicialmente es $F_{in} = 0, 2$, tiene un máximo $F_p = 1, 0 = 100 \text{ pN}$ menor que F_M y un valor final $F_f = 0, 1 = 10 \text{ pN}$ menor que F_m .

En la figura 6 se muestra la extensión escalonada de la proteína, que ha sido sometida a una fuerza $F_p = 1 = 100 \text{ pN}$, con $F_{c2} < F_p < F_M$. Cada escalón en la extensión de la proteína se corresponde con la extensión de un módulo, que van procediendo ordenadamente del primero hasta el último. El pliegue de la proteína es similar al de la figura 5.

El estado inicial necesario para que se produzca este tipo de extensión en los experimentos, con el primer módulo extendido, se puede obtener de varias formas: bien porque la proteína se encontraba en ese estado debido a una manipulación anterior, porque se ha extendido debido al ruido térmico o porque los parámetros del primer módulo son diferentes al resto y tiene una fuerza umbral para la extensión F'_M , de forma que $F'_M < F_p < F_M$, y se extiende directamente. Un ejemplo de este último caso es que el primer módulo tenga una longitud $L'_c > L_c$, de forma que su fuerza umbral es $F'_M = F_M L_c / L'_c$.

4. Estimación del tiempo

Para estimar el tiempo característico de la extensión de cada módulo nos fijamos en el segundo módulo cuando se aumenta bruscamente la fuerza hasta F_p , que se encuentra en el intervalo $F_{c2} < F_p < F_M$. El primer módulo tiene una extensión de $u^{(3)}(F_{c2})$, los módulos con $j > 2$ tienen una extensión $u^{(1)}(F_{c2})$ y el segundo módulo \tilde{u} , ligeramente mayor. Para la fuerza crítica F_{c2} se cumple que $\kappa[u^{(1)}(F_{c2}) + u^{(3)}(F_{c2})] = V'(\tilde{u}) + 2\kappa$, y $2\kappa + V''(\tilde{u}) = 0$. Si tomamos la ecuación (6) para $j = 2$ y la expandemos alrededor de \tilde{u} , siguiendo las ideas de [1], se tiene

$$\begin{aligned} \dot{u}_2 &\sim F_p - F_{c2} - V'(\tilde{u}) - V''(\tilde{u})(u_2 - \tilde{u}) - \frac{V'''(\tilde{u})}{2}(u_2 - \tilde{u})^2 + \kappa[u^{(1)}(F_{c2}) + u^{(3)}(F_{c2})] \\ &\Rightarrow \dot{u}_2 \sim F_p - F_{c2} - \frac{V'''(\tilde{u})}{2}(u_2 - \tilde{u})^2. \end{aligned} \quad (10)$$

Como κ es pequeño entonces \tilde{u} es muy cercano a u_M , y $V'''(\tilde{u}) \simeq V'''(u_M) < 0$. Con esta aproximación, la ecuación diferencial (10) tiene la solución

$$u_2 = \tilde{u} + \frac{\sqrt{2(F_p - F_{c2})|V'''(u_M)|}}{|V'''(u_M)|} \tan\left(\sqrt{\frac{(F_p - F_{c2})|V'''(u_M)|}{2}}(t - t_2)\right), \quad (11)$$

donde t_2 es una constante. Inicialmente, en $t = t_2$, la solución es $u_2 = \tilde{u}$ y cuando t llega al valor donde $\sqrt{(F_p - F_{c2})|V'''(u_M)|}/2 (t - t_2) = \pi/2$ la solución se va a infinito. En ese momento en el que se produce un salto es cuando el segundo módulo llega a $u^{(3)}(F_{c2})$. Por lo tanto el tiempo característico de este salto es $\pi/\sqrt{2(F_p - F_{c2})|V'''(u_M)|}$. Después de que el segundo módulo se haya extendido, el tercer módulo tendría una extensión \tilde{u} y se encontraría en la misma situación. Este proceso se repetiría con todos los módulos hasta llegar al último, tras el cual la proteína estaría completamente extendida, produciendo la curva escalonada característica de la figura 6.

Para el proceso de extensión simultáneo que hemos visto antes se puede hacer un razonamiento similar, donde todos los módulos se extienden a la vez en un tiempo característico $\pi/\sqrt{2(F_p - F_M)|V'''(u_M)|}$. El proceso de plegado es paralelo, con un tiempo de característico de $\pi/\sqrt{2(F_m - F_f)|V'''(u_m)|}$. Como la tercera derivada del

potencial es mucho mayor en valor absoluto en u_m que en u_M el tiempo de plegado es mayor que el de extendido, como veíamos en la figura 5.

V. RESULTADOS NUMÉRICOS PARA LA DINÁMICA ESTOCÁSTICA

Ahora consideramos el mismo estudio considerando también la parte estocástica de la ecuación (6). Al resolver una realización hay una componente de ruido blanco, haciendo que no sea necesario sobrepasar los umbrales de fuerza para que se produzca la extensión y el pliegue de los módulos. Para la extensión, la fuerza F_p puede ser menor que F_M y se puede producir la extensión de un módulo, sin necesidad de estar en la configuración inicial que se necesitaba en la extensión escalonada en la parte determinista. Para el pliegue ocurre lo mismo, pudiendo ser F_f mayor que F_m para que los módulos lleguen al mínimo entálpico. Estos procesos ocurren aleatoriamente, permaneciendo un tiempo en un estado metaestable y llegando a la configuración estable, que siempre ocurre para un tiempo suficientemente grande. El tiempo medio en el que el módulo se encuentra en el estado metaestable depende de la diferencia entre F_M y F_p para la extensión, y F_f y F_m para la contracción, haciéndose menor cuanto menor es la diferencia. Como la amplitud del ruido depende de la temperatura a la que se encuentra el sistema también influye en los tiempos de estos procesos.

En la figura 7 se muestra el plegado de los módulos de la proteína modular, afectado por el ruido térmico del sistema a una temperatura $T = 300 K$. Se observa claramente que el tiempo que transcurre en el proceso de plegado en (a) que tiene una fuerza $F_f = 0,14$ más cercana a $F_m = 0,104$ es relativamente pequeño, $t_p \simeq 200$. En (b) que tiene una fuerza $F_f = 0,16$, mayor, el tiempo de plegado es mucho más grande, $t_p \simeq 900$.

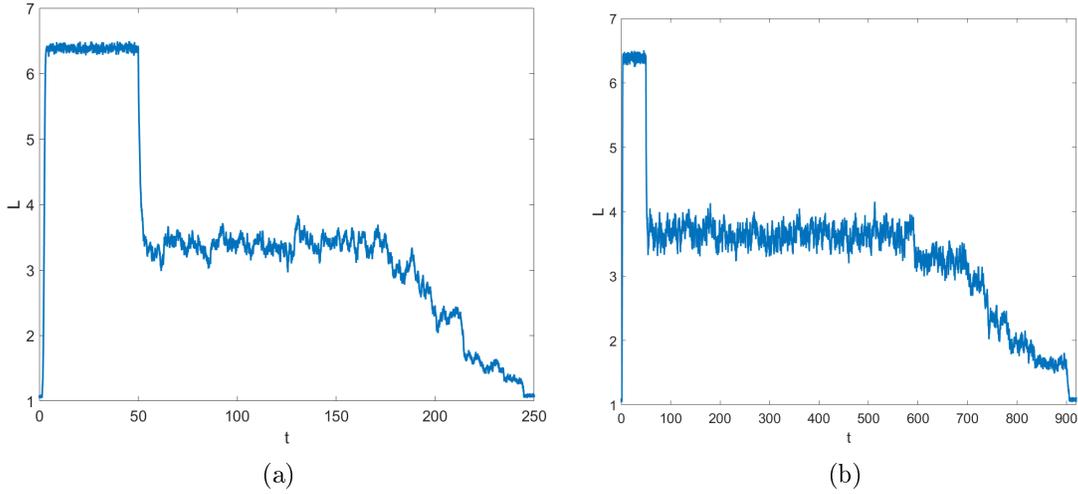


Figura 7: Curva de extensión de la proteína sometida a una fuerzas $F_{in} = 0$, $F_p = 1$, (a) $F_f = 0,14$ (b) $F_f = 0,16$. Puede verse el efecto del ruido, con una constante de amplitud $\theta = 0,0014$. Se3 puede observar que el tiempo máximo necesario para representar todo el proceso en (b), donde la fuerza F_f es bastante mayor que F_m , es mucho mayor que el de (a) donde la proteína se pliega completamente más rápido.

En la figura 8 se grafican dos extensiones escalonadas con los mismos parámetros, resolviendo la ecuación de Langevin con ruido para fuerzas $F_{in} = 0,2$, $F_p = 0,62 < F_M$ y $F_f = 1,16 > F_m$. Aparece la misma estructura escalonada en la extensión como en la figura 6, pero aquí los tiempos entre extensión de diferentes módulos son aleatorios y en general mucho mayores. El ruido interfiere con el mecanismo que veíamos en la extensión escalonada determinista, pudiendo tener tiempos en general menores o mayores en función de la temperatura. En el proceso de plegado la proteína que disminuye su longitud hasta llegar a la extensión $u^{(3)}(F_f)$ y permanece ahí un tiempo aleatorio, hasta que por efecto del ruido un módulo se pliega, con una extensión $u^{(1)}(F_f)$. Desde esa configuración el resto de módulos se van plegando en un tiempo relativamente corto hasta llegar al mínimo entálpico en todos. Otra diferencia con la figura 6 es que en la realización estocástica no es necesario que la proteína se encuentre en el estado inicial concreto mencionado anteriormente para la extensión escalonada sin ruido, produciéndose la extensión siempre con un tiempo suficientemente grande debido al ruido.

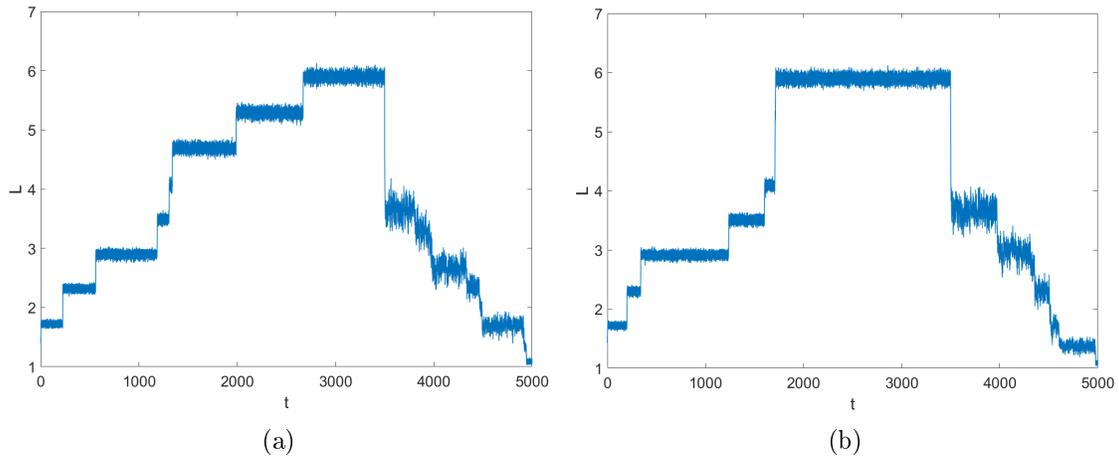


Figura 8: Dos realizaciones de la extensión y pliegue de la proteína sometida a fuerzas $F_{in} = 0,2$, $F_p = 0,62$, $F_f = 0,16$, manteniendo cada una hasta que se llega al estado de equilibrio. Puede verse que en ambas realizaciones los estados inicial, intermedio y final son iguales, pero los tiempos a los que se extiende cada módulo y se pliegan son diferentes en cada una.

Los resultados obtenidos de la expansión escalonada aleatoria son similares a los obtenidos en los experimentos en [5].

VI. CONCLUSIONES

En este trabajo hemos visto un modelo teórico para el estudio de las propiedades elásticas de proteínas modulares en experimentos de fuerza-extensión y extensión-abrazadera. Con este modelo hemos podido hacer simulaciones, a través de integración numérica en códigos de MATLAB, que concuerdan con los resultados experimentales y se observan características en la extensión y pliegue de proteínas reales. En la simulación de la curva de fuerza-extensión se obtiene la curva característica de dientes de sierra y se explica en términos de fuerzas críticas y dinámica con un potencial. Con el mismo modelo se ha obtenido la extensión simultánea de la proteína para fuerzas mayores que la necesaria para la transición al estado de extensión, y la extensión escalonada para fuerzas menores pero mayores que una determinada fuerza crítica. Se ha estudiado el mecanismo de extensión de los módulos en este último caso y el tiempo en el que transcurre, a partir de frentes de onda y estados virtuales. Por último se ha estudiado el efecto del ruido que es introducido al sistema por la temperatura en el proceso de extensión y plegado, y cómo modifica los tiempos característicos de estos procesos. En el trabajo se han aplicado métodos y técnicas aprendidas en diferentes asignaturas del Máster.

VII. CÓDIGOS MATLAB

1. Resolución numérica de la curva fuerza-extensión

Todos los códigos se han implementado y ejecutado en MATLAB R2017a.

```
% Script para calcular la curva fuerza-extensión de una
% proteína modular
% Diego Botija 2018
clear all; clc;

Fp=100;
P=0.28;
Lc=30;
U0=200;
Rc=4;
b=2;
T=300;
kb=0.013806;
D=1500;
eta=kb*T/D;

global mu beta rho kappa A theta Fin Lfij;

mu=U0/(Lc*Fp);
beta=2*b*Lc/Rc;
rho=Rc/Lc;
kappa=0.065;
A=kb*T*Lc/(4*P*U0);
theta=kb*T/(Fp*Lc);

N=8;
T=300;
dt=0.05;
t=dt:dt:T;
% u0=[u3Fin*[1,1],u1Fin*ones(1,N)];
u0=0.12*[ones(N+2,1)];

for i=1:numel(t)
    Lfij=1+4*t(i)*0.005;
    Li(i)=Lfij;
    u=fminsearch(@(u)potencialMin(u),u0);
    u0=u;
    Fi(i)=1/N*(-kappa*(u(1)-u(2)-u(N+1)+u(N+2)));
    for l=1:N
        Fi(i)=Fi(i)+1/N*dPotencial(u(1+l));
    end
end
plot(Li,Fi,'b','LineWidth',2.0)
hold on

for i=1:numel(t)
```

```

Lfi j=7-4*t(i)*0.005;
Ld(i)=Lfi j;
%v0=u0(2:N+1);
u=fminsearch(@(u)potencialMin(u),u0);
u0=u;
Fd(i)=1/N*(-kappa*(u(1)-u(2)-u(N+1)+u(N+2)));
for l=1:N
    Fd(i)=Fd(i)+1/N*dPotencial(u(1+l));
end
end
plot(Ld,Fd,'r','LineWidth',2.0)
xlabel('L')
ylabel('F')

```

```

% Derivada del potencial adimensionalizado, V'(u)
function dV = dPotencial(u)

global mu beta rho kappa A theta;
dV=mu*(2*(1-exp(-beta*(u-rho)))*beta.*exp(-beta*(u-rho))+...
    A*(1./(1-u).^2-1+4*u));

end

```

```

% Energía libre total de la proteína, con términos parabólicos para
% forzar las restricciones
function A = potencialMin(u)

global kappa Lfi j;
A=0;
for i=2:numel(u)-1
    A=A+PotProt(u(i));
end
for i=3:numel(u)-1
    A=A+kappa/2*(u(i)-u(i-1))^2;
end
A=A+50*(sum(u(2:numel(u)-1))-Lfi j)^2+20*(u(1)-u(2))^2+...
    20*(u(numel(u))-u(numel(u)-1))^2;

end

```

```

% Potencial adimensional de un módulo de la proteína
function V = PotProt(u)

global mu beta rho kappa A theta;
V=mu*((1-exp(-beta*(u-rho))).^2-1+A*(1./(1-u)-1-u+2*u.^2));

end

```

2. Resolución numérica de la dinámica determinista

```
% Programa para calcular la extensión L(t) de una proteína modular
% sometida a una fuerza F(t), para la expresión determinista de la ...
% ec. Langevin
% Diego Botija 2018
clear all;clc;

Fp=100;
P=0.28;
Lc=30;
U0=200;
Rc=4;
b=2;
Temp=300;
kb=0.013806;
D=1500;
eta=kb*Temp/D;

global mu beta rho kappa A theta F Fin;

mu=U0/(Lc*Fp);
beta=2*b*Lc/Rc;
rho=Rc/Lc;
kappa=0.065;
A=kb*Temp*Lc/(4*P*U0);
theta=kb*Temp/(Fp*Lc);

Fin=0.2;
u1Fin=fzero(@dPotencialF,0.1);
u3Fin=fzero(@dPotencialF,0.8);

N=8;
T=50;
%u0=u1Fin*[ones(1,N+2)];
u0=[u3Fin*[1,1],u1Fin*ones(1,N)];

% Resuelto con ode23t
[t,u]=ode23t(@Langevin,[0 T],u0);
L1=sum(u(:,2:9)');
figure(1)
plot(t,L1,'LineWidth',2.0)
axis([0 T 1 7])
xlabel('t'); ylabel('L'); title('L vs t (ode23t)')

% Resuelto con Euler explícito
dt=0.001;
t=0:dt:T;
u=u0;
L2(1)=sum(u(2:N+1));
F=Fin;
for j=2:numel(t)
```

```

    if t(j)>1.5 && t(j)<27
        F=1.0;
    end
    if t(j)>27
        F=0.1;
    end
    for i=2:N+1
        un(i)=u(i)+dt*(F-dPotencial(u(i))+kappa*(u(i-1)+u(i+1)-2*u(i)));
    end
    un(1)=un(2);
    un(N+2)=un(N+1);
    u=un;
    L2(j)=sum(u(2:N+1));
end
figure(2)
plot(t,L2,'LineWidth',2.0)
axis([0 T 1 7])
xlabel('t'); ylabel('L'); title('L vs t (Euler explícito)')

```

```

% Término de la derecha de la ec. de Langevin sin ruido
function dudt = Langevin(t,u)

global F kappa;
F=0.2;
if t>1.5 && t<27
    F=1.0;
end
if t≥27
    F=0.1;
end

N=numel(u);
for i=2:N-1
    dudt(i)=F-dPotencial(u(i))+kappa*(u(i-1)+u(i+1)-2*u(i));
end
dudt(1)=dudt(2);
dudt(N)=dudt(N-1);
dudt=dudt';

end

```

```

% Derivada del potencial menos la fuerza aplicada, V'(u)-F
function dV = dPotencialF(u)
global mu beta rho kappa A theta Fin;

dV=mu*(2*(1-exp(-beta*(u-rho)))*beta.*exp(-beta*(u-rho))+...
    A*(1./(1-u).^2-1+4*u))-Fin;
end

```

3. Resolución numérica de la dinámica estocástica

```
% Programa para calcular la extensión L(t) de una proteína modular
% sometida a una fuerza F(t), para la expresión determinista de la ...
% ec. Langevin
% Diego Botija 2018
clear all;clc;

Fp=100;
P=0.28;
Lc=30;
U0=200;
Rc=4;
b=2;
Temp=300;
kb=0.013806;
D=1500;
eta=kb*Temp/D;

global mu beta rho kappa A theta F Fin;

mu=U0/(Lc*Fp);
beta=2*b*Lc/Rc;
rho=Rc/Lc;
kappa=0.065;
A=kb*Temp*Lc/(4*P*U0);
theta=kb*Temp/(Fp*Lc);

Fin=0.0;
u1Fin=fzero(@dPotencialF,0.1);
u3Fin=fzero(@dPotencialF,0.8);

N=8;
T=250;
%u0=u1Fin*[ones(1,N+2)];
u0=[u3Fin*[1,1],u1Fin*ones(1,N)];

% Resuelto con Euler-Mayurama
dt=0.01;
t=0:dt:T;
u=u0;
L2(1)=sum(u(2:N+1));
F=Fin;
for j=2:numel(t)
    if t(j)>1.5 && t(j)<50
        F=1.0;
    end
    if t(j)>=50
        F=0.14;
    end
    for i=2:N+1
        un(i)=u(i)+dt*(F-dPotencial(u(i))+kappa*(u(i-1))+...
```

```

                u(i+1)-2*u(i))+sqrt(dt*2*theta)*randn;
    end
    un(1)=un(2);
    un(N+2)=un(N+1);
    u=un;
    L2(j)=sum(u(2:N+1));
end
figure(2)
plot(t,L2,'LineWidth',2.0)
axis([0 T 1 7])
xlabel('t'); ylabel('L'); title('L vs t (Euler-Maruyama)')

```

Bibliografía

- [1] L. L. Bonilla, A. Carpio and A. Prados, *Protein unfolding and refolding as transitions through virtual states*, EPL, **108** (2014) 28002.
- [2] R. Berkovich et al., *Collapse Dynamics of Single Proteins Extended by Force*, Biophys. J., **98** (2010) 2692.
- [3] J. F. Marko and E. D. Siggia, *Statistical mechanics of supercoiled DNA*, Phys. Rev. E **52**, (1995) 2912.
- [4] T. E. Fisher, A. F. Oberhauser, M. Carrion-Vazquez, P. E. Marszalek, and J. M. Fernandez, *The study of protein mechanics with the atomic force microscope*, Trends Biochem. Sci. **24**, 379 (1999).
- [5] Sergi Garcia-Manyes, Jasna Brujić, Carmen L. Badilla, and Julio M. Fernández, *Force-Clamp Spectroscopy of Single-Protein Monomers Reveals the Individual Unfolding and Folding Pathways of I27 and Ubiquitin*, Biophys. J., **93** (2007) 2436.
- [6] Kloeden, P.E. and Platen, E. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin. (1992) ISBN 3-540-54062-8.