

Reducción de Pasos de Tiempo en Trayectorias de Plegamiento de Proteínas

19 de junio de 2017

1. Resumen

El análisis de las trayectorias de dinámica molecular para proteínas es de gran importancia ya que nos muestra como la proteína cambia durante su proceso de plegamiento. Generalmente, las simulaciones muestran como una proteína inicia en un estado no-plegado y a través de distintas transformaciones durante el tiempo de su plegamiento, esta alcanza un estado más estable (*estado nativo*). En este trabajo implementamos distintas técnicas para reducir los pasos de tiempo de los datos provenientes de las simulación con dinámica molecular del plegamiento de proteínas. Los resultados de las distintas técnicas se comparan mediante dos métodos: uno donde se calculan y muestran los puntos críticos de las trayectorias resultantes; y otro donde se calcula la dimensión fractal de las trayectorias. El objetivo al final es obtener la trayectoria reducida más representativa de tal manera que el número de pasos de tiempo se haya reducido considerablemente y por lo tanto los análisis se puedan llevar a cabo más rápidamente.

2. Introducción

2.1. Estado del Arte

2.1.1. Trayectorias de Proteínas

1. Trayectorias por Dinámica Molecular
 - a) Background de que son y cuales son sus características
 - b) Mención de las trayectorias públicas disponibles: caso de Anton (12 proteínas) y caso de folding@home (proteína villin)
2. Trayectorias por otros métodos (desplegamiento)
 - Como se realizan y características
 - Caso de las trayectorias generadas por el servidor Parasol de la U de Texas, profe. Amato

2.1.2. Reducción de la dimensionalidad para trayectorias de proteínas

Corto análisis de lo que presentan los siguientes artículos y que tienen o que no tienen frente a las técnicas que planteamos. Esencialmente, estos trabajos reducen la dimensionalidad capturando características relevantes de las trayectorias pero sin buscar preservar la forma en que se lleva el plegamiento, es decir, obtienen los momentos más relevantes mientras que nosotros buscamos que la trayectoria reducida conserve en lo posible los eventos que la caracterizan tales como los momentos donde la proteína se pliega y se vuelve a desplegar (mínimos y máximos en los puntos críticos).

- Rajan, A., Freddolino, P. L., & Schulten, K. (2010). Going beyond clustering in MD trajectory analysis: An application to villin headpiece folding. PLoS ONE, 5(4), e9890. <http://doi.org/10.1371/journal.pone.0009890>
- Eitrich, T., Mohanty, S., Xiao, X., & Hansmann, U. H. E. (2007). Dimensionality Reduction Techniques for Protein Folding Trajectories. From Computational Biophysics to Systems Biology (CBSB07), 36, 99–102.

- Plaku, E., Stamati, H., Clementi, C., & Kavraki, L. E. (2007). Fast and reliable analysis of molecular motion using proximity relations and dimensionality reduction. *Proteins: Structure, Function, and Bioinformatics*, 67(4), 897–907.

3. Datos y Métodos

3.1. Trayectorias simuladas de plegamiento de proteínas

- Las CARACTERÍSTICAS de las trayectorias y proteínas que utilizamos para realizar las reducciones. corresponden a las trayectorias puestas al público por el grupo del Dr David Shaw (Anton). Completar la información sobre estas trayectorias, creo que algo había en mi tesis de grado o sino de los paper originales o sino de otros papers que hayan utilizado estas trayectorias y den una breve descripción de estos conjunto. Datos que importan es tipo de simulación, time step, máquina o servidor, proteínas simuladas. Todo con sus referencias principalmente hacia los papers de David Shaw donde presenta las simulaciones.
- El primer caso es el la 2JFO, toca hablar de esta. Y si existe un segundo caso después se agrega.

3.2. Reducciones

La forma de reducir las trayectorias se realiza a travez de particionamientos con un número de conformaciones que corresponden al orden que se busca reducir, es decir una reducción de *nanos* a *picos* (en orden de 1000), dividirá la trayectoria en particiones de 1000 conformaciones a partir de la cual se seleccionará una conformación representativa de acuerdo a una de las siguientes técnicas: agrupamiento alrededor de medoides; agrupamiento por jerarquías; selección estática, y selección aleatoria.

El agrupamiento alrededor de medoides particiona los datos en k grupos y selecciona de ese grupo el *medoide* o conformación más cercana a todos las demás conformaciones. Para nuestro caso, la partición corresponde al orden de la reducción que se quiera realizar, es decir si se va a reducir de *nanos* a *picos* (en orden de 1000), entonces se divide la trayectoria en particiones de 1000 conformaciones y cada partición corresponde a uno de los k grupos de donde se selecciona el medoide. Al final los medoides seleccionados corresponderán a la nueva trayectoria reducida.

El agrupamiento jerarquico alrededor de grupos,

Finalmente, el agrupamiento estático selecciona la conformación de la mitad del pedazo de trayectoria que corresponde a la partición. Y el agrupamiento aleatorio selecciona una conformación al azar dentro de las conformaciones de la partición. En ambos métodos, la selección de la conformación no tiene en cuenta que tan próxima o lejana esta está de las demás.

3.3. Validación

3.3.1. Validación por Puntos Críticos

Hablar sobre el que se pretende y como se lleva el cálculo. Por ahora es el único que he visto como medio para comprobar que tan buena es una reducción Lo que se busca con este método es encontrar los puntos críticos en la trayectoria principal (máximos y mínimos) y por cada trayectoria reducida encontrarle los puntos críticos y compararlos, al final debe resultar un conteo que se podría mostrar en una simple tabla

	Máximos	Mínimos	Diferencia frente a la trayectoria completa
Medoides			
Jerarquico			
Estatico			
Aleatorio			

Punto crítico (matemáticas)

Para otros usos de este término, véase [Punto crítico](#).

En [cálculo](#), un **punto crítico** de una [función](#) de una [variable real](#) es cualquier valor en el [dominio](#) en [diferenciable](#) o cuando su [derivada](#) es 0.^{1 2} El valor de la función en el punto crítico es un **valor crítico**. Las definiciones admiten generalizaciones a funciones de varias variables, mapas diferenciables entre [R](#) entre [variedades diferenciables](#).

4. Resultados

4.1. Reducción

A continuación mostramos los gráficos de las trayectorias resultantes después de la reducción. El primer caso se realizó para la trayectoria de la proteína con código PDB 2JFO (*ENTEROCOCCUS FAECALIS GLUTAMATE RACEMASE IN COMPLEX WITH D- AND L-GLUTAMATE*). Las trayectorias que se muestran son: la trayectoria completa en la Fig. a; por agrupamiento kmeans y selección de medoides en la Fig. b; por agrupamiento jerárquico y selección de representativos en la Fig. c; por selección estática en la Fig. d; y por selección aleatoria en la Fig. e. La línea negra muestra la trayectoria reducida mientras que la línea roja muestra la trayectoria de forma suavizada. En general se observa que la forma de la trayectoria se conserva en todas las reducciones.

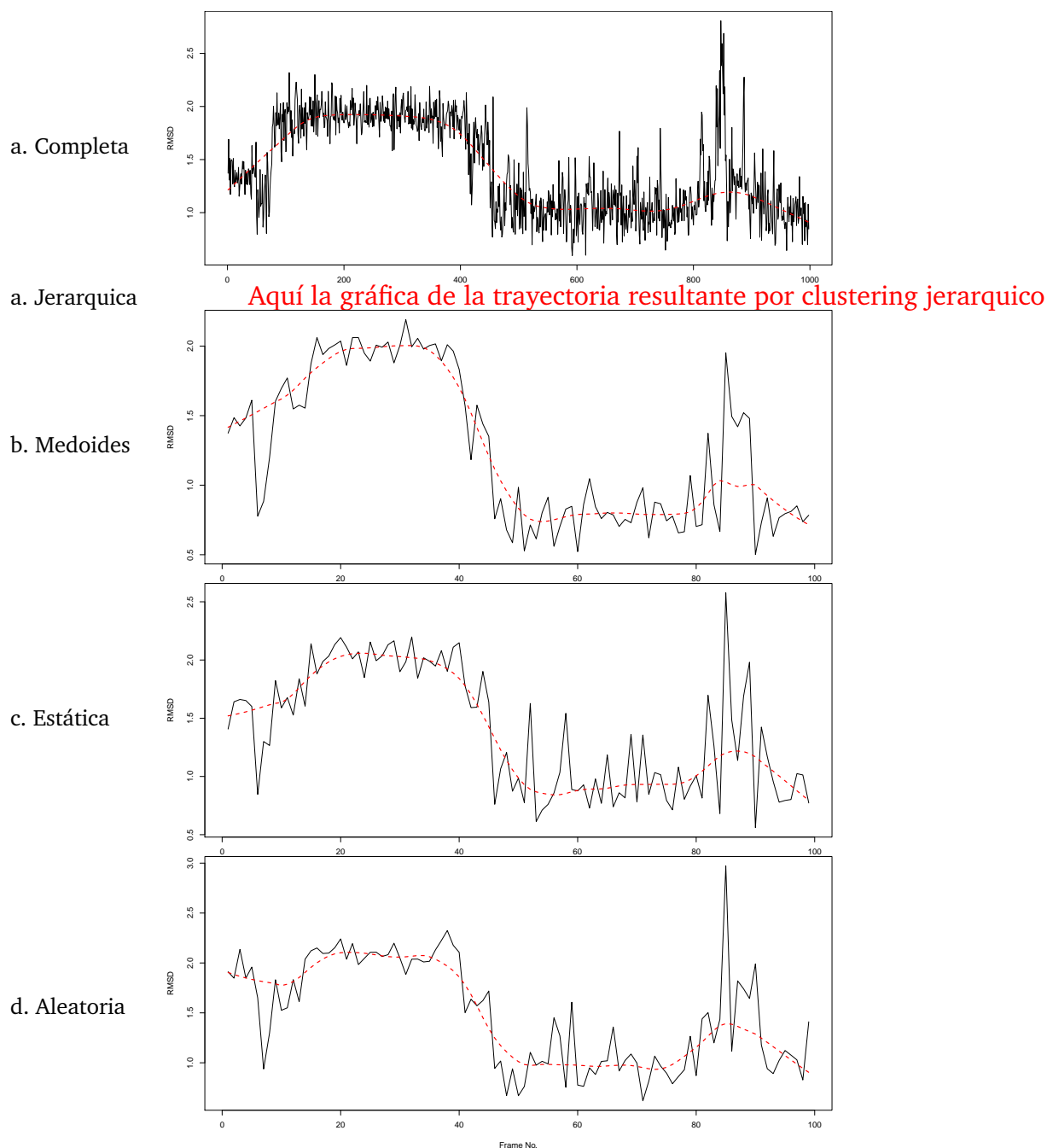


Figura 1: Trayectoria de plegamiento de la Proteína 2JOF y cuatro trayectorias reducidas. Las reducciones corresponde al orden de un 1ns es decir 1000 pasos se reducen a 1 solo paso.

4.2. Análisis

Aquí va el cálculo de los puntos críticos que debe implementarse. Yo encontré una función en R que creo hace todo ese cálculo y creo que necesita como argumentos los valores de donde salen las gráficas, es decir, el nombre de la proteína y su valor RMSD que debería calcularse cuando se hace la reducción.

5. Referencias

Las referencias se deberían manejar en algún manejador, mejor dicho utilizar MendeleyDesktop, yo ya tengo varias de estas sino todas y a medida que aparezca yo puedo adicionarla a Mendeley y tener compartido el repositorio.

La forma de pegarlas aca es exportarla a un archivo BibTex y adicionarlo, así

Referencias