

Algoritmo para Simplificar Trayectorias de Plegamiento de Proteínas

Luis Garreta¹, Mauricio Martinez², Pedro A. Moreno³

{luis.garreta, mauricio.martinez.jimenez, pedro.moreno}@correounivalle.edu.co

Grupo de Investigación en Bioinformática
Universidad del Valle
Cali, Colombia

Resumen

Las simulaciones del proceso de plegamiento de las proteínas han demostrado ser de gran utilidad para estudiar las mecanismos subyacentes de este proceso, el cual permite a una cadena de aminoácidos plegarse hasta alcanzar su estructura tridimensional única y convertirse en una proteína activa habilitada para ejecutar una función biológica. Gracias a los avances en hardware y software, estas simulaciones han experimentado grandes progresos, con tiempos de simulación en el orden de los milisegundos, y llevadas a cabo usando diferentes tecnologías, desde costosas supercomputadoras especializadas [4], hasta arreglos de tarjetas gráficas [5], e incluso PCs distribuidos alrededor del mundo [2]. Muchas de estas simulaciones alcanzan tiempos que antes no se lograban debido a las limitaciones en el hardware, y las trayectorias que generan abarcan miles o millones de conformaciones, lo cual a pesar de ser una gran ventaja porque se tiene más detalle del proceso, así mismo es un problema debido al tiempo y recursos computacionales necesarios para analizarlas. Por esta razón, se necesitan nuevos algoritmos capaces de simplificar estas trayectorias de una forma rápida, y que logren preservar la mayor información posible tanto a nivel de representación como a nivel de orden temporal de las conformaciones de la trayectoria.

Para realizar estas reducciones se han utilizado diferentes técnicas que básicamente caen en dos enfoques: la reducción de dimensionalidad [1] y el agrupamiento [7], las cuales más que simplificar las trayectorias realizan un análisis sobre ellas. En el primer enfoque se transforma las conformaciones a una forma simplificada para poder interpretar los resultados, sin embargo, aunque se conserva el orden temporal de las conformaciones, su estructura se pierde ya que su representación se modifica. En cambio, en el segundo, se encuentran grupos de conformaciones con propiedades comunes donde las conformaciones conservan su estructura, pero el orden temporal se pierde ya que los grupos pueden contener conformaciones de tiempos muy diferentes.

En este trabajo presentamos un algoritmo para simplificar trayectorias de plegamiento de proteínas que divide una trayectoria de plegamiento en segmentos y por cada segmento extrae rápidamente las conformaciones más disimilares, tomando como base la estrategia rápida de agrupamiento de Hobohm&Sander [3], y luego selecciona de forma más exhaustiva a las más representativas, mediante una estrategia de particionamiento alrededor de medoides [6]. El algoritmo aprovecha el orden temporal implícito en la trayectoria para realizar en cada segmento comparaciones locales y evitar la comparación de todos contra todos, que se vuelve impráctica computacionalmente cuando son muchas conformaciones. De esta manera, el algoritmo reduce muy rápidamente la trayectoria y las conformaciones seleccionadas conservan tanto su estructura como su orden temporal. Además, la partición por segmentos permite al algoritmo simplificar cada segmento de forma independiente y paralela, lo que lo vuelve aún más rápido cuando se ejecuta en máquinas multi-núcleo, muy comunes hoy en día.

Referencias

- [1] Mojie Duan, Jue Fan, Minghai Li, Li Han, and Shuanghong Huo. Evaluation of Dimensionality-reduction Methods from Peptide Folding-unfolding Simulations. *Journal of chemical theory and computation*, 9(5):2490–2497, may 2013.
- [2] Daniel L Ensign, Peter M Kasson, and Vijay S Pande. Heterogeneity even at the speed limit of folding : Large-scale molecular dynamics study of a fast-folding variant of the villin headpiece. *Journal of molecular biology*, 374(3):806–816, 2007.
- [3] Uwe Hobohm, Michael Scharf, Reinhard Schneider, and Chris Sander. Selection of representative protein data sets. *Protein Science*, 1(3):409–417, mar 1992.
- [4] Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, and David E Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, oct 2011.
- [5] Hai Nguyen, James Maier, He Huang, Victoria Perrone, and Carlos Simmerling. Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. *Journal of the American Chemical Society*, 136(40):13959–13962, oct 2014.
- [6] Hae-Sang Park and Chi-Hyuck Jun. A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications*, 36(2):3336–3341, mar 2009.
- [7] Jun-hui Peng, Wei Wang, Ye-qing Yu, Han-lin Gu, and Xuhui Huang. Clustering algorithms to analyze molecular dynamics simulation trajectories for complex chemical and biological systems. *Chinese Journal of Chemical Physics*, 31(4):404–420, aug 2018.