



Temática: Bioinformática— Matemática computacional

Modelo de representación de superficies de estructuras macromoleculares

Surface representation model of macromolecular structures

Edisel Navas-Conyedo^{1*}, Leonardo A. Ramírez-Cayón², Amauris D. González-Suárez², Jorge Gulín-González¹

¹Centro de Estudio de Matemática Computacional (CEMC). Grupo de Matemática y Física Computacionales, Facultad de Ciencias y Tecnologías Computacionales. Universidad de las Ciencias Informáticas, Carretera a San Antonio delos Baños, 2 $\frac{1}{2}$ Km, Torrens, La Lisa, La Habana, Cuba.

²Facultad de Ciencias y Tecnologías Computacionales. Universidad de las Ciencias Informáticas, Carretera a San Antonio delos Baños, 2 $\frac{1}{2}$ Km, Torrens, La Lisa, La Habana, Cuba.

*Autor para correspondencia: enavas@uci.cu

Resumen

Algunas estructuras biológicas como macromoléculas, proteínas y plaquetas tienen cientos o miles de estructuras atómicas, donde las estructuras internas son irrelevantes a la hora de evaluar la interacción entre ellas. Al evaluar la interacción entre estructuras, la información presente en su superficie es suficiente, por esta razón, es importante contar con un modelo que abstraiga toda la estructura interna, manteniendo solo la presente en la superficie y evitando cálculos innecesarios. Proponemos un módulo de Python nombrado `pysurmolmesh` para la construcción de un modelo representativo de las superficies de estructuras macromoleculares a través de una malla de partículas interactuantes que reflejan sus propiedades mecánicas y químicas, evitando contener información sobre las estructuras macromoleculares internas.

Palabras claves: Modelo de superficie, Macromoléculas, Python

Abstract

Some biological structures as macromolecules, proteins, and platelets have hundreds to thousands of atomic structures, where internal structures are irrelevant when evaluating the interaction between them. When evaluating the interaction between structures, the information present on their surface is sufficient, for this reason, it is important to have a model that abstracts the entire internal structure, keeping only the one present on the surface and avoiding unnecessary computations. We propose a Python module named `pysurmolmesh` for the construction of a representative model of the surfaces of macromolecular structures through a mesh of interacting particles that reflect their mechanical and chemical properties, avoiding containing information on the internal macromolecular structures.

Keywords: *Surface model, Macromolecule, Python*



Introducción

Para el estudio computacional de sistemas macromoleculares con métodos ab-initio como dinámica molecular (Boinepalli and Attard, 2003) o similares, se necesita de la información que especifica las características morfológicas e interacciones del sistema de estudio. Usualmente, el usuario en el proceso de diseño de la simulación debe proveer explícitamente toda la información de todos los componentes y las interacciones presentes en el sistema, donde en algunos casos existen comportamientos inalterables e irrelevantes para el estudio realizado (LAM, 2021). Tal es el caso de las interacciones entre sistemas macromoleculares, puesto que estas ocurren solamente entre las estructuras que conforman la superficie, quedando inmutable el resto de las estructuras internas. Esto posibilita la creación de un modelo abstracto que reduce el volumen de información correspondiente a las estructuras con conformaciones internas inmutables, reduciendo la información a un modelo de superficie.

En general, las proteínas tienen características cercanas a cuerpos semi-rígidos, que pueden modelarse como una colección de partículas coloides unidas (Fuhrmans and Müller, 2013; Kreuziger et al., 2020). Para que se tenga una idea cuantitativa, estructuras como la Trombina poseen 9278 átomos y de ellos solo 456 en la capa superficial (Chinnaraj et al., 2018), por tanto, con un modelo de 456 partículas o menos se pueden representar las características químico-físicas de esta molécula durante su interacción con otros componentes. Desde el punto de vista físico, estas estructuras pueden considerarse materia condensada suave (*soft condensed matter en Inglés*), donde las propiedades elásticas del enmallado y los potenciales iónicos-químicos definen su comportamiento, siendo relevante construir el modelo físico acorde a parámetros experimentales por interpolación por análisis posteriores Ab-initio (Bagchi et al., 2005; Grubmüller et al., 1991; von Bülow et al., 2019). De esta manera se preserva el efecto de forma y tamaño de la estructura, evitando la contribución de las estructuras internas, manteniendo la sola información necesaria para evaluar la interacción con otras estructuras durante la simulación. La herramienta de software para la generación del modelo de superficie es implementada como un módulo para *Python* (Foundation, 2020), optimizado para su uso en el ambiente interactivo Jupyter (Varoquaux et al., 2021), el cual incluye además de las funciones necesarias para generar el modelo de superficie, otras útiles para su posterior representación visual 3D.

Materiales y métodos o Metodología computacional

Una macromolécula está compuesta por un conjunto de N átomos de los cuáles se conoce su disposición espacial, además de cada átomo se conoce su tamaño efectivo, para lo cual se puede usar el radio iónico como referencia. Por tanto, de cada i -átomo de la estructura se conoce la tupla $\{\vec{R}_i, d_i\}$, siendo \vec{R}_i el vector de posición de cada átomo y d_i su radio iónico. Para la reconstrucción de la superficie exterior de la estructura es necesario comprender el concepto de función de distancia signada de un punto p en el espacio respecto a una



esfera E con centro en la posición \vec{R}_E y de radio d_E , como:

$$s_{pE} = \left| \vec{r}_p - \vec{R}_E \right| - d_E \quad (1)$$

Resulta evidente comprobar en la ecuación (1) que para los puntos interiores a la esfera $s_{pE} < 0$, para los exteriores $s_{pE} > 0$ y $s_{pE} = 0$ en los puntos sobre la superficie. Por tanto, la distancia de un punto p a la superficie externa de una estructura formada por los N átomos se puede calcular como

$$s_{pN} = \text{mín}(s_{pi}) \quad (2)$$

, siendo s_{pi} la distancia del punto p al i -átomo de la estructura (Persson and Strang, 2004; Zhang et al., 2014).

Para la generación del modelo de superficie externas primeramente se genera un conjunto de puntos equidistantes sobre una esfera que circunscribe a la estructura, centrada en $\vec{R}_C = \frac{1}{N} \sum_i^N \vec{R}_i$ de radio $d = \text{máx} \left(\left| \vec{R}_i - \vec{R}_C \right| \right) + \text{máx}(d_i)$. El número de puntos depende del número de latitudes $\rho \geq 2$, generando puntos sobre la superficie esférica que equidistan en

$$s = 2 \pi d \sin \left(\frac{\pi}{4\rho} \right) \quad (3)$$

con un número de puntos aproximado de $m = 5\rho^2 - \rho + 2$ ¹. Cada punto está conectado con sus vecinos más cercanos formando un un enmallado de Delaunay (Varoquaux et al., 2021). En la Figura 1 se muestran la distribución de puntos generadas para el conjunto de valores $\rho \in \{2, 4, 5\}$ correspondientes a $\{20, 78, 124\}$ puntos respectivamente.

Cada punto j generado es movido hacia la superficie de la estructura bajo la acción de una pseudo-fuerza correspondiente a un potencial elástico $U_j = \frac{1}{2} \left(s_{jN}^2 + \sum_{jk} \left(\left| \vec{R}_j - \vec{R}_k \right| - s \right)^2 \right)$, donde el índice k corresponde a los vecinos de la triangulación que conecta cada vértice, el proceso se realiza a través del método de gradiente descendiente (Lambert, 1991; Galassi et al., 2011; Press et al., 2007). Este proceso garantiza que se mantenga la distancia media entre los puntos de la superficie generada a la vez que se sitúen sobre la superficie exterior con una densidad constante que atrapa las características geométricas fundamentales de la estructura y reduce el volumen de puntos a generar.

Para acelerar la evaluación de la distancia en la ecuación (2), se almacena la información de la posición de los N átomos de la estructura en una estructura de datos de árbol KD (KD Tree en idioma Inglés) (Varoquaux et al.,

¹Realmente $m = \text{floor} (5,14 \rho^2 - 1,6\rho + 2)$ obtenido por ajuste polinomial de segundo orden.

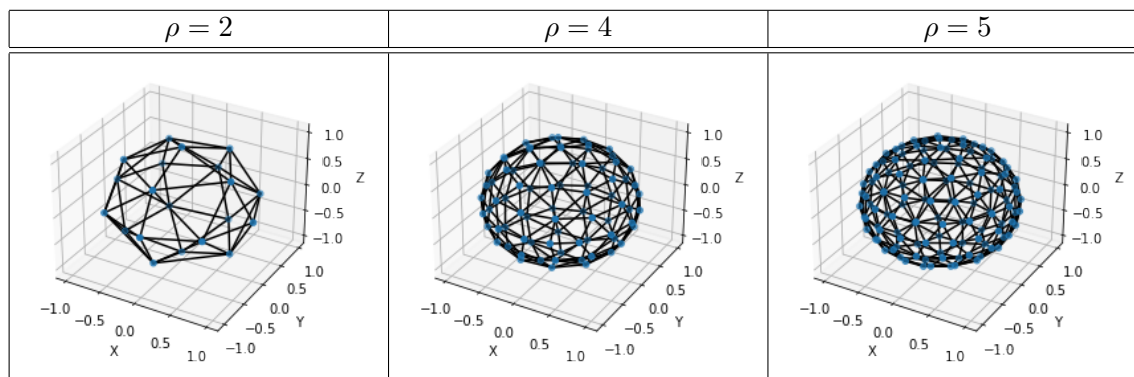


Figura 1: Distribución de puntos en una esfera unitaria para $\rho \in \{2, 4, 5\}$.

2021), con lo que se asegura que la evaluación tenga un orden de complejidad algorítmica de $\mathcal{O}(\log_2(N))$. En el enmallado de Delaunay 3D se genera un conjunto de tetraedros de los cuáles se eliminan las caras interiores, evaluando la función de distancia signada (2) sobre los incentros de los triángulos de cada cara, dejando solo la cara de menor valor absoluto de distancia.

El algoritmo para generar el modelo de superficie ha sido implementado como un módulo de Python optimizado para su uso desde interfaces interactivas como Jupyter, el cual se puede instalar utilizando la herramienta de gestión de paquetes de Python *pip* desde la URL <https://github.com/Ertytux/pysurmolmesh.git> a través de:

```
pip install -e https://github.com/Ertytux/pysurmolmesh.git
```

Para cargar una estructura macromolecular, desde un archivo en alguno de los formatos estándares para este tipo de estructura, se utiliza la función *loadPDB* que devuelve tres arreglos, el primero de ellos el listado de tipos de átomos, el segundo sus posiciones y por último el radio iónico de cada tipo como:

```
import pysurmolmesh.loadPDB as lpd

filename="/path/to/file.pdb"

atypes , apos , aradi=lpd.loadPDB(filename)
```

generando luego la superficie:

```
import pysurmolmesh.proctools as mpt
```



```
vsi=mpt.VesicleGenerator (apos , aradi)
```

```
rho=4
```

```
pts , bl , trigs , err=vsi.generator (rho)
```

El parámetro ρ es el ρ utilizado para generar el enmallado inicial relajado sobre la estructura, el cuál define la finura utilizada y se escoge buscando que el número de vértices sea menor o igual que el número de átomos exteriores de la estructura, a la vez que se tenga la información geométrica necesaria para construir un modelo reducido. Los parámetros de salida son: *pst* es el arreglo de posiciones de los vértices del enmallado, *bl* el arreglo de los pares de índices de las aristas, *trigs* el arreglo de la triada de vértices de los triángulos de las caras de las superficies exteriores y por último *err* el residuo relativo al valor s dependiente de ρ como medida del error de generación, ver ecuación (3).

Resultados y discusión

Para ilustrar la representación simplificada de estructuras, se utilizaron tres macromoléculas con geometrías diferentes: Trombina (Chinnaraj et al., 2018), Amiliod-P y proteína C-Reactiva (Mikolajek et al., 2011), estas estructuras tienen en común que poseen forma propelar, por lo que resulta compleja su representación adecuada, pues poseen bordes agudos con ángulos de ataque progresivos, aspecto que es importante para muchos estudios donde la hidrodinámica es importante (Xu et al., 2020; BaÅYaÄYaoÄYlu et al., 2018). Para cada una de ellas hicimos una representación simple de las estructuras atómicas, donde por simplicidad y para una representación visual adecuada mostramos las posiciones atómicas de los átomos de mayor tamaño. A continuación generamos el correspondiente enmallado y luego mostramos conjuntamente el enmallado y las representaciones atómicas, el cual denominamos mixto.

La Trombina, Figura 2, tiene una estructura alargada con un centro ancho, semejante a una hélice, donde los puntos geométricos relevantes se encuentran en el centro y bordes de las propelas, este tipo de estructura es relevante pues los bordes deben identificarse y hay una estructura progresiva que es muy particular de las formas propelares. Esto es identificado por el algoritmo concentrando un mayor número de puntos en el centro y borde de la estructura.

Tanto el Amiliod-P, Figura 3, como la proteína C-Reactiva, Figura 4, poseen una estructura pentagonal, donde la diferencia más notable está en el grosor casi cilíndrico del Amiliod-P con una cavidad interna más pequeña.

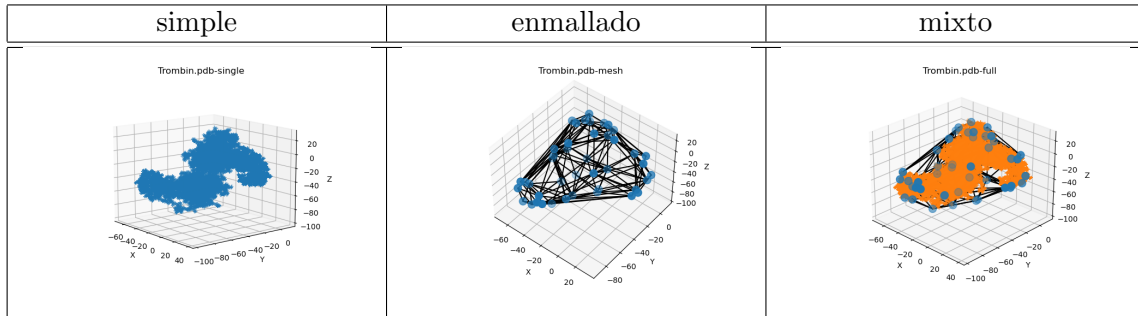


Figura 2: Representación de la Trombina

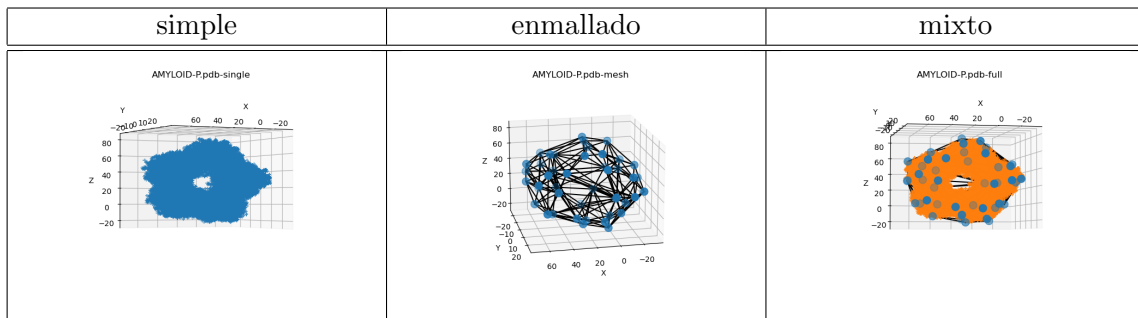


Figura 3: Representación del Amilode-P

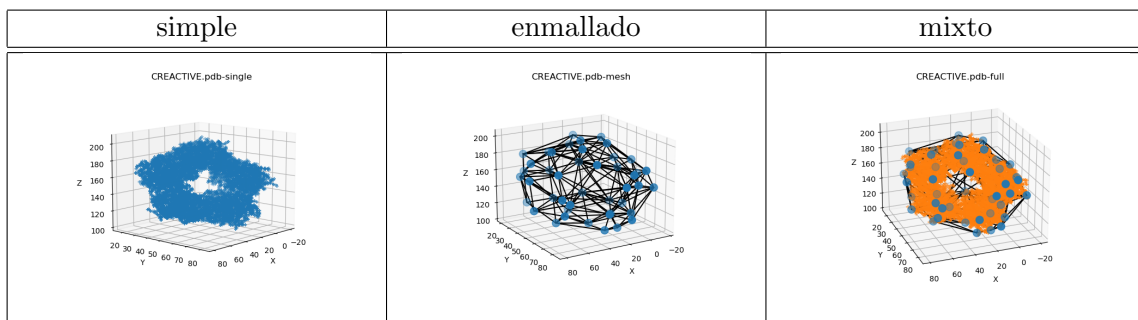


Figura 4: Representación de la proteína C-Reactiva



En ambos casos se logra una representación adecuada de las estructuras externas, no así de la cavidad, como en el Amiliodo-P, donde los puntos del enmallado generado cerca de la cavidad interior no son suficientes, aún cuando la estructura general es aceptable.

Cuando el enmallado es generado, además de la información de los vértices, se genera la información de las conexiones entre ellos y de la lista de triadas de puntos que conforman las superficies de las caras. Esta información resulta útil para generar estructuras vesiculares coloidales que sustituyan a la macromolécula durante una simulación. Téngase en cuenta que en las interacciones entre estructuras macromoleculares esencialmente intervienen las superficies, la parte interna solo se tienen en cuenta cuando por motivos eléctricos (desplazamiento de cargas o efectos de dipolos eléctricos) intervienen en las interacciones, aunque este aspecto se puede atribuir directamente a los nodos situados en superficie de manera efectiva, de forma que se reproduzca aproximadamente el mismo patrón de interacción.

En general, la herramienta desarrollada genera correctamente el modelo de superficie de las estructuras cuando no hay cavidades internas dentro de la estructura a representar, esto constituye una limitante si se desea estudiar efectos o fenómenos de cavitación o absorción de estructuras más pequeñas dentro de la cavidad, lo cual se será corregido en versiones futuras del módulo.

Conclusiones

Se desarrolló un módulo de Python para su uso interactivo desde Júpiter para creación de modelos de superficie y la representación geométrica de estructuras macromoleculares. El algoritmo utilizado para generar el modelo de superficie reproduce las características geométricas fundamentales de la macromolécula representada, salvo cuando hay presencia de cavidades, donde correcciones posteriores deben ser introducidas.

La información del enmallado generado puede asociarse a una estructura coloidal sustituyente para ser utilizado en simulaciones complejas, donde se evalúe solo la interacción entre las superficies de los componentes o en estudios de microfluídica con la interacción con un solvente.

Agradecimientos

Agradecemos al proyecto nacional e institucional de la Universidad de las Ciencias Informáticas: *Estudio de los efectos hidrodinámicos en las interacciones de partículas virales en la sangre mediante técnicas de microfluídica computacional* por el soporte de la presente investigación.



Referencias

- (2021). *LAMMPS Users Manual*. Sandia National Laboratories, <http://lammps.sandia.gov>.
- Bagchi, P., Johnson, P. C., and Popel, A. S. (2005). Computational fluid dynamic simulation of aggregation of deformable cells in a shear flow. *Journal of Biomechanical Engineering*, 127(7):1070–1080.
- BaÅÿaÄÿaoÄÿlu, H., Succi, S., Wyrick, D., and Blount, J. (2018). Particle shape influences settling and sorting behavior in microfluidic domains. *Scientific Reports*, 8(1):8583.
- Boinepalli, S. and Attard, P. (2003). Grand canonical molecular dynamics. *J. Chem. Phys.*, 119:12769–12775.
- Chinnaraj, M., Z.Chen, Pelc, L. A., Grese, Z., Bystranowska, D., Cera, E., and Pozzi, N. (2018). Structure of prothrombin in the closed form reveals new details on the mechanism of activation. *SCI REP*, 8(2945).
- Foundation, P. S. (2020). Python 3.8.10. Electronic Manual.
- Fuhrmans, M. and Müller, M. (2013). Mechanisms of vesicle spreading on surfaces: coarse-grained simulations. *Langmuir*, 29:4335–4349.
- Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Booth, M., and Rossi, F. (2011). *Gnu Scientific Library: Reference Manual*. Network Theory Ltd.
- Grubmüller, H., Heller, H., Windemuth, A., and Schulten, K. (1991). Generalized verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation*, 6(1-3):121–142.
- Kreuziger, L. B., Lee, A., David Garcia, A. C., Cushman, M., DeSancho, M., and Connors, J. M. (2020). Covid-19 and vte/anticoagulation: Frequently asked questions.
- Lambert, J. D. (1991). *Numerical Methods for Ordinary Differential Systems*. John Wiley and Sons, Chichester.
- Mikolajek, H., Kolstoe, S. E., Pye, V. E., Mangione, P., Pepys, M. B., and Wood, S. P. (2011). Structural basis of ligand specificity in the human pentraxins, c-reactive protein and serum amyloid p component. *Journal of Molecular Recognition*, 24(2):371–377.
- Persson, P.-O. and Strang, G. (2004). A simple mesh generator in matlab. *SIAM Review*, 46(2):329–345.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical recipes in C++*. *The art of scientific computing*. Cambridge: University Press, —c2007, Third Edition.
- Varoquaux, G., Gouillart, E., Vahtras, O., de Buyl, P., and many others (2021). *Scipy Lecture Notes*.



von Bülow, S., Siggel, M., Linke, M., and Hummer, G. (2019). Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. *Proceedings of the National Academy of Sciences*, 116(20):9843–9852.

Xu, P., Zhou, Q., and Xu, J. (2020). Mechanism of thrombocytopenia in covid-19 patients. *Annals of Hematology*.

Zhang, N., Zhang, P., Kang, W., Bluestein, D., and Deng, Y. (2014). Parameterizing the morse potential for coarse-grained modeling of blood plasma. *Journal of Computational Physics*, 257, Part A:726–736.