

Proteína E del SARS-CoV-2 y amantadina

David De la Mora González

8 de enero de 2021

Abstract

El presente trabajo contiene información sobre la proteína E del SARS-CoV-2 y de la amantadina, una molécula pequeña que podría usarse como tratamiento del CoViD-19. Como parte del trabajo se incluyen los resultados obtenidos con un script de Python que analiza el genoma del SARS-CoV-2 y es capaz de encontrar la proteína E en él. Posteriormente, se realiza un análisis de acoplamiento molecular con el programa AutoDock Tools de ambas moléculas y se obtienen las conformaciones y energías de unión de los acoplamientos. Al final, se obtiene un archivo pdb con la estructura tridimensional del mejor acoplamiento obtenido.

1 Introducción

El SARS-CoV-2 es un coronavirus causante de la enfermedad conocida como CoViD-19, que a la fecha ha causado cerca de 88.2 millones de contagios y 1.9 millones de muertes en todo el mundo, y en México ha afectado a 1.49 millones de habitantes y resultado en el fallecimiento de cerca de 131 000 personas.

El objetivo del presente trabajo se centra en la proteína E del SARS-CoV-2. En primer lugar, se diseñó un programa en Python para encontrar la secuencia del gen y de la proteína en el contexto del genoma completo del virus, y en segundo lugar se procedió a realizar un análisis de acoplamiento molecular con la amantadina (AMT), un medicamento antiviral.

1.1 SARS-CoV-2

El SARS-CoV-2 se notificó inicialmente el 31 de diciembre de 2019 en pacientes infectados en la ciudad de Wuhan, China. Este virus de RNA posee envoltura, es de cadena sencilla y sentido, con una longitud genómica de 29 903 bp de acuerdo a la secuencia publicada con el número de acceso NC_045512 en la base de datos del NCBI. El periodo de incubación promedio antes de desarrollar síntomas se calcula que es de 5 días. El genoma codifica para 5 genes esenciales que codifican para 4 proteínas estructurales. El orden de los genes en el genoma es el siguiente: 5'-ORF de la replicasa 1b-S-E-M-N-3'. El genoma se encuentra envuelto con la nucleoproteína N. En la nucleocápside se encuentran las proteínas spike (S), de membrana (M), y de envoltura (E)[1], en la cual se enfocará este trabajo.

Se sabe que el receptor del virus en humanos que permite la entrada viral a través de la proteína S es la enzima convertidora de angiotensina (ACE) 2, la cual se expresa principalmente en células alveolares en el pulmón. La respuesta inflamatoria a la infección es muy exacerbada e implica una "tormenta de citocinas" que contribuyen a la disfunción multiorgánica, al igual que la replicación viral descontrolada y la respuesta inflamatoria perjudicial[1].

1.1.1 Proteína E

La proteína E del SARS-CoV-2 es una viriporina de 75 aminoácidos que forma un canal selectivo para cationes a lo largo de la membrana viral [2]. Existe en la base de datos de proteína (PDB) la estructura cristalográfica de esta estructura y corresponde al código de acceso 7K3G, en la cual se observa que la proteína se ensambla en la membrana como un homopentámero. Esta estructura consiste únicamente en la región transmembranal que consiste en los aminoácidos 8-38. El diámetro del canal varía de 11-14 Å. El poro del canal está ocupado predominantemente por residuos hidrofóbicos, incluyendo Asn15, Leu18, Leu21,

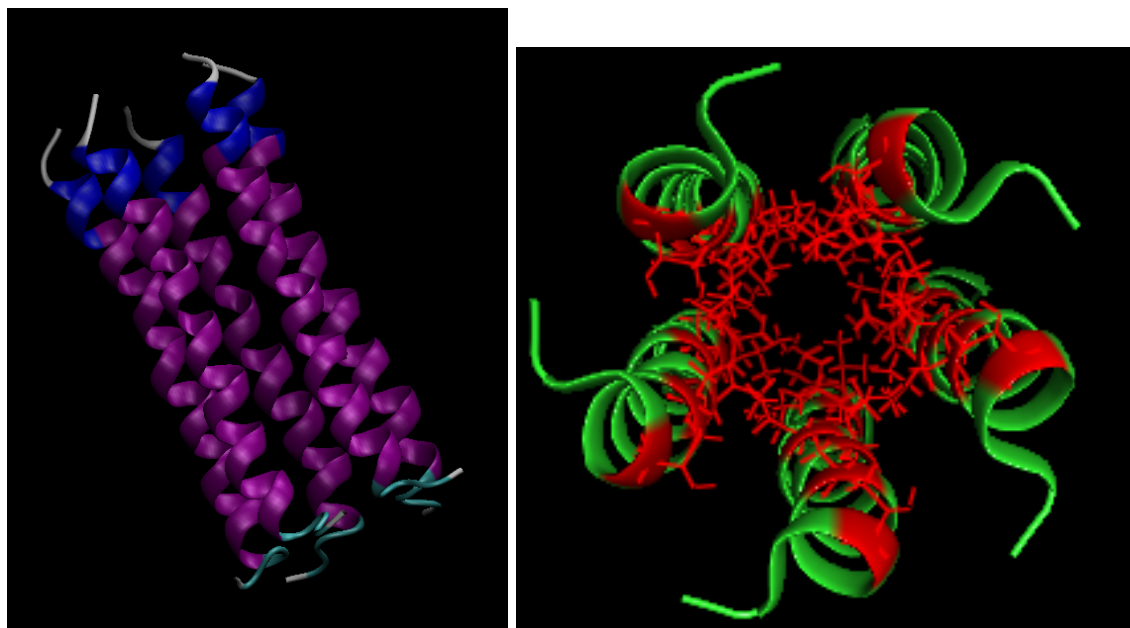


Figure 1: *Izquierda:* Se muestra la estructura del pdb del pentámero de la proteína E. *Derecha:* Vista superior donde se señalan de rojo los aminoácidos que forman parte del poro.

Val25, Leu28, Ala32 y Thr35. Sin embargo, se cree que esta estructura representa el estado cerrado del poro, y en realidad el estado abierto podría consistir en un poco más grande e hidratado [2].

1.2 Amantadina

La amantadina es un fármaco antiviral que comúnmente se usa para el tratamiento de la influenza A. Su fórmula molecular es $C_{10}H_{17}N$ y tiene un peso de 151.25 g/mol. En la base de datos del DrugBank tiene un número de acceso de DB00915.

Se sabe que la AMT bloquea la actividad del canal formado por la proteína E uniéndose cerca de la región N-terminal, por lo que puede ser entonces un fármaco antiviral candidato contra el SARS-CoV-2 [2].

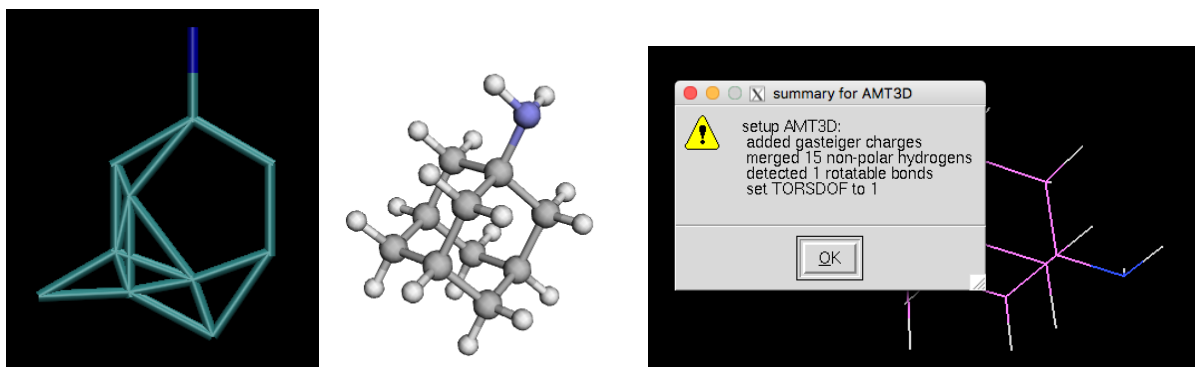


Figure 2: *Izquierda:* Estructura plana de la AMT donde se muestra la conectividad de los C en azul claro y del N en azul oscuro. *Centro:* Estructura tridimensional que muestra todos los átomos de la AMT. *Derecha:* Mensaje que aparece en el AutoDock Tools al cargar la molécula en 3D

```

Secuencia genómica y de aminoácidos de la proteína E:
AUGUACUCAUUCGUUUCGGAAGAGACAGGUACGUUAAUAGUAAUAGCGUACUUCUUUUUCUUGCUUUCGUGGUAUUCUUGCUAGUUAACACUAGCCAUCUUACUGCGCUUCGAUUGUGGCGUACU
GCGCAAAUUAUUGUAAACGUGAGUCUUGUAAACCUUCUUUUUACGUUUUACUCUCUGUGUAAAAUUCUGAAUUCUUCUAGAGUCCUGAUCUUCUGGUC 225

MYSFVSEETGLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVLPSTFYVYSRVNLNSSRPDLLV 75
>YP_009724392.1 envelope protein [Severe acute respiratory syndrome coronavirus 2]
MYSFVSEETGLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKLNLSRV
PDLV

```

Figure 3: *Arriba*: Resultado de la ejecución del script en Python, que muestra la secuencia genómica y de aminoácidos de la proteína E encontrada en el genoma del SARS-CoV-2. *Abajo*: Secuencia de referencia de la proteína E del SARS-CoV-2 obtenida de la base de datos del NCBI.

2 Resultados

El procedimiento para realizar este trabajo consistió en los siguiente pasos:

1. Se ejecutó el script de Python que analiza el genoma del SARS-CoV-2 para encontrar la proteína E y se compara con la secuencia en formato fasta de la proteína E en la base de datos de la NCBI. Los resultados se observan en la Figura 3.
2. Se cargaron en el programa de AutoDock Tools las estructuras tridimensionales tanto de la proteína E del SARS-CoV-2 como de la AMT. En la Figura 4 se observa la zona de trabajo con las moléculas cargadas.
3. Se prepararon los archivos de entrada para poder realizar el análisis del acoplamiento molecular. Las imágenes y archivos de entrada y de salida se adjunta de manera suplementaria.
4. Ya que se sabe que la AMT se acopla cerca del extremo N-terminal, se seleccionó una cuadrícula cerca de esta zona, tal y como se observa en la Figura 5.
5. Tras ejecutar el acoplamiento molecular con los parámetros preestablecidos, se resumen en la Tabla 1 los resultados de los diez acoplamientos moleculares obtenidos.
6. Finalmente, del primer acoplamiento se obtuvo la estructura tridimensional en formato pdb que se muestra en la Figura 6

2.1 Localización de la proteína E del SARS-CoV-2 en su secuencia genómica utilizando un script de Python

Para este resultado, se realizó un script en Python que es capaz de leer un archivo de entrada .txt que posee la secuencia genómica del SARS-CoV-2 y buscar en los tres posibles marcos abiertos de lectura de la cadena sentido posibles inicios y finales de la transcripción génica. Una vez localizados, el script también es capaz de leer los codones y traducirlos al aminoácido correspondiente y así obtener tanto la secuencia genómica como protéica. Como se sabe que la proteína E que buscamos consiste en 75 residuos, se estableció un filtro en el script para que únicamente se muestren los resultados correspondientes a una secuencia de 75 aminoácidos, con lo que se obtuvo como resultado lo que se muestra en la Figura 3. Esto se comparó con la secuencia de referencia de la proteína E que se encuentra en la base de datos del NCBI y, como se puede observar, sí corresponde de manera idéntica lo encontrado con la ejecución del script con la secuencia de referencia.

2.2 Acoplamiento molecular en AutoDock Tools

El archivo tridimensional de la proteína E del SARS-CoV-2 que se encuentra disponible en la base de datos de PDB, consiste en un homopentámero. Cabe resaltar que no corresponde a la proteína completa, sino que consiste únicamente en la región transmembranal. Este archivo contiene 10 conformeros diferentes, de los

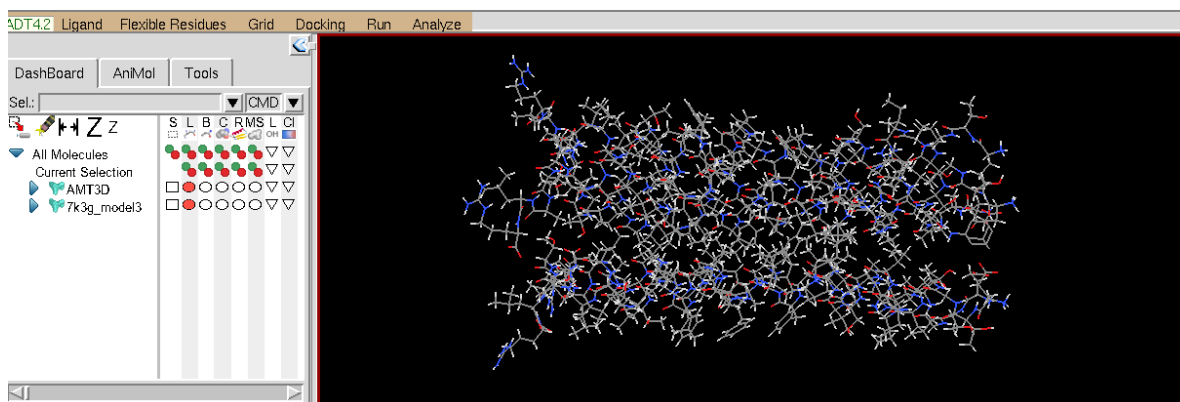


Figure 4: Se muestra el área de trabajo de AutoDock Tools con las moléculas tridimensionales de la proteína E del SARS-CoV-2 y la AMT

Acoplamiento	Energía de Acoplamiento (kcal/mol)	Constante de Afinidad (μ M)
1	-4.89	261.68
2	-4.87	267.27
3	-4.87	267.36
4	-4.87	267.52
5	-4.87	271.47
6	-4.85	280.86
7	-4.83	286.01
8	-4.83	289.77
9	-4.82	291.23
10	-4.82	293.30

Table 1: Resultado de las energías de acoplamiento y la constante de afinidad calculada para los diez acoplamientos obtenidos de AutoDock Tools de la proteína E con la AMT.

cuales se seleccionó al azar el conformero 3 para realizar el acoplamiento. Es posible observar que los 10 son diferentes entre sí, por lo que el único criterio tomado en cuenta fue una selección por números aleatorios del 1-10.

Ya se ha reportado que la AMT se puede acoplar a la proteína E del SARS-CoV-2 en la región N-terminal, es por ello que la cuadrícula seleccionada para el análisis se encuentra en esta zona y los parámetros, tanto de tamaño como de ubicación, se observan en la Figura 5.

De manera suplementaria se adjuntan capturas de pantalla de todos los parámetros seleccionados para el análisis, de los cuales la mayoría son los que se encuentran por defecto. También se adjuntan los archivos de entrada y de salida generados. Al realizar el acoplamiento molecular, se obtuvieron 10 posibles resultados que se muestran en la Tabla 1. En esta tabla se anota la energía de acoplamiento y la constante de afinidad calculada. Al respecto, cabe mencionar que el hecho de que los valores sean negativos implican que esta reacción de unión/acoplamiento es espontánea y termodinámicamente favorable. Los valores son todos muy cercanos entre sí, y las imágenes adjuntas de cada uno de los acoplamientos muestran que no existe mucha diferencia entre ellos, lo cual se refleja en los valores cercanos de la energía de acoplamiento. Sin embargo, es posible observar cambios más sutiles en la afinidad al analizar los resultados de la constante de afinidad. Cabe recordar que, entre menor o más pequeña sea, implica que hay una mayor afinidad. Es por ello que el acoplamiento 1 es la conformación que muestra una mayor afinidad y, el acoplamiento 10, la menor afinidad de los 10 mostrados.

Finalmente, se seleccionó el acoplamiento 1 por presentar los mejores valores calculados de acoplamiento para generar una estructura tridimensional en pdb de la proteína E del SARS-CoV-2 acoplada a la AMT, como se observa en la Figura 6.

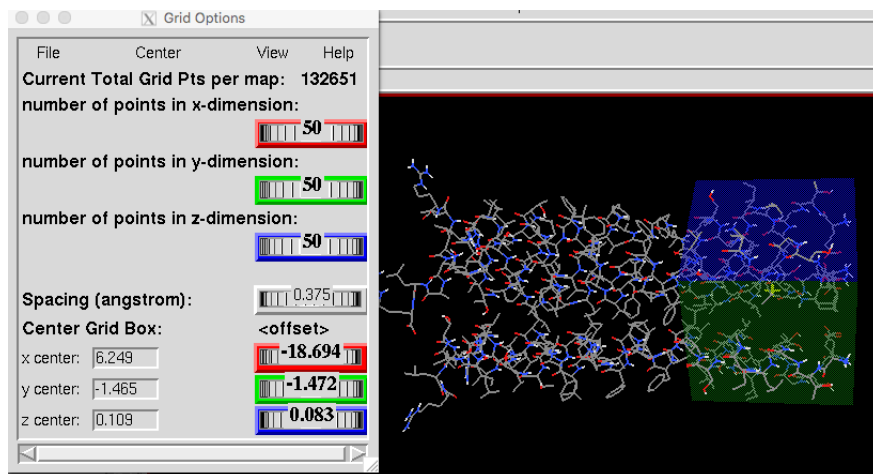


Figure 5: Se muestran los parámetros seleccionados para la realización de la cuadrícula para el análisis del acoplamiento molecular, la cual se ubicó en el extremo N-terminal

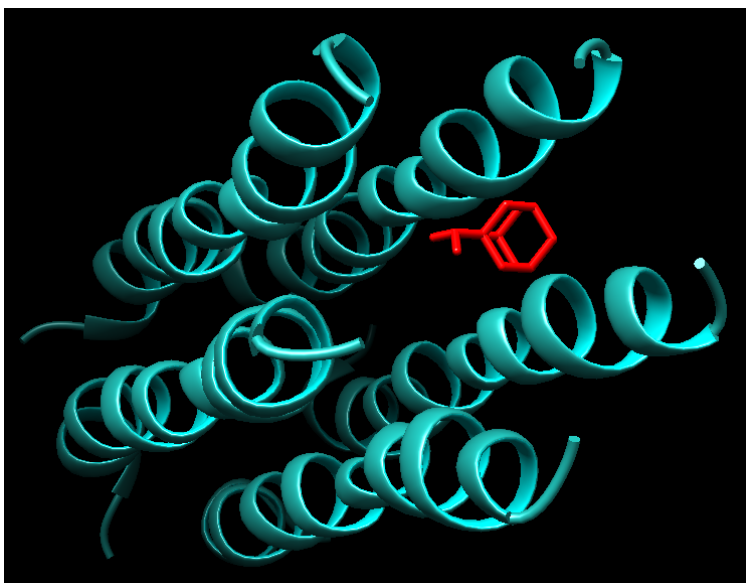


Figure 6: Estructura en formato pdb del acoplamiento molecular 1 obtenido tras el análisis en AutoDock Tools. La proteína E del SARS-CoV-2 se muestra en azul y la AMT en rojo

3 Conclusiones

Este trabajo muestra que el script realizado en Python es capaz de encontrar la proteína E del SARS-CoV-2 a partir de su secuencia genómica completa y que corresponde totalmente con la secuencia de referencia. También fue posible realizar un análisis de acoplamiento molecular del pentámero formado por esta proteína E con el fármaco antiviral AMT y donde se corrobora que existe un acoplamiento entre ambas moléculas. Este primer acercamiento teórico requeriría posteriores experimentos *in vitro*, posteriormente, *in vivo*, para evaluar la relevancia farmacológica de la AMT y su posible función en el bloqueo de este canal como posible estrategia para combatir la enfermedad CoViD-19.

References

- [1] Neil A. Bhowmick et al. “COVID-19 and androgen-targeted therapy for prostate cancer patients”. In: *Endocrine-Related Cancer* 27.9 (2020), R281–R292. ISSN: 14796821. DOI: 10.1530/ERC-20-0165.
- [2] Venkata S. Mandala et al. “Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers”. In: *Nature Structural and Molecular Biology* 27.December (2020). ISSN: 15459985. DOI: 10.1038/s41594-020-00536-8. URL: <http://dx.doi.org/10.1038/s41594-020-00536-8>.