

Análisis de la Nsp15 del SARS-COV2: una aproximación bioinformática en la lucha contra la COVID-19

Juan Manuel Ruiz Robles

20 de febrero de 2021



UNIVERSIDAD DE GRANADA

Índice

1. Análisis genético	3
2. Nsp15: Non Structural Protein 15	8
2.1. Dominio NendoU	13
2.2. Dominio intermedio	15
2.3. Dominio de oligomerización	16
2.4. Hexámero	19
3. Opciones terapéuticas	21
3.1. Saikosaponinas	22
3.2. Tipiracil	24
3.3. Ciclenosida	25
4. Actividades	27
4.1. Bases de datos secuenciales y estructurales	27
4.1.1. Diagrama de flujo	27
4.1.2. Cargar PDB	29
4.2. Visualización de proteínas	31
4.2.1. Ficheros de estructuras	31
4.2.2. WritePDB	31
4.3. Predicción estructural de proteínas	33
4.3.1. Ramachandran	33
4.3.2. Coordenadas transformadas	36
4.3.3. Alinear Z	38
4.3.4. RMSD	41
4.3.5. Mutación	42
4.3.6. Enlaces disulfuro	46
4.3.7. Anfipatía-hidrofobicidad	48
Referencias	51

1. Análisis genético

En primer lugar y como complemento al trabajo, he decidido aportar mi toque personal con un breve análisis bioinformática de la secuencia del SARS-COV2. En NCBI descargué la secuencia completa del virus en formato FASTA, típicamente utilizado en el ámbito científico.

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC_045512.2

[GenBank](#) [Graphics](#)

```
>NC_045512.2 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome
ATTAAGGTTTACCTTCCCAGTAACAAACCAACCAACTTCGATCTTGAGATCTGCTCAA
CGACTTTAAATCTGTGGCTGCACTCGCTGCATGCTTAGTGCACTCACGAGTATAATTAAAC
TAATTACTGTCGTTGACAGGACACGAGTAACCGCTATCTCTGAGGCTGCTACGTTTCGCGTG
TTGCAAGCCGATCATCAGCACATCTAGGTTCTCGCCGGTTGACCGAAAGGTAAGATGGAGAGCCTTGTG
CCTGGTTCAACGAGAAAACACAGTCCAACCTCAGTTGCCCTGTTTACAGGTTCCGCGACGTGCTCGTAC
GTGGCTTGGAGACTCGTGGAGGAGCTTATCAGAGGCCAGTCACATCTAAAGATGGACTTGTGG
CTTAGTAGAAGTTGAAAAAGGCGTTTGCTCAACTTGAACAGCCCTATGTTCATCAAACGTTGGAT
```

Esto lo hice porque tengo varias funciones desarrolladas en el ordenador para el lenguaje de programación Python, las cuales me permiten detectar ciertas características de las secuencias genéticas, lo que en este caso, debido a la relevancia de la biología funcional en el estudio del SARS-COV2 tiene cabida en este trabajo.

Todas las funciones utilizadas son de elaboración propia, realizadas a partir de autoaprendizaje visitando videotutoriales y páginas web. Dichas funciones tratan de imitar los procesos bioquímicos básicos de la célula.

En general, mi forma de trabajo consiste en tomar la/las secuencias e introducirlas en un archivo FASTA, y tengo una función que me crea una estructura de diccionario en Python, donde el nombre de la secuencia es la clave del diccionario y la secuencia es el “value” o elemento correspondiente a la clave.

```
1 # -*- coding: utf-8 -*-
2
3 # Created on Tue Jan 19 12:42:12 2021
4
5 @author: Juanna
6
7
8 from bio_seq import bio_seq
9 from DNA_structures import writeTextFile, readTextFile, read_FASTA
10
11 fasta= read_FASTA("text.txt")
12 # print(fasta)
13 for f in fasta:
14
15     print ( "Este es el nombre de la cadena: " + f)
16     print(f)
17     print ( "Este es la secuencia: " + fasta[f])
18
```

De esta forma, podemos tener toda secuencia identificada mientras estudiamos diversas características de la misma.

Por ejemplo, podemos observar la longitud de la secuencia:

The screenshot shows a Jupyter Notebook environment. On the left, the code cell contains Python code for reading a FASTA file and creating a sequence object. The right side shows the terminal output where the sequence is printed as a long string of nucleotides, followed by its type, label, and length.

```
1 # -*- coding: utf-8 -*-
2 
3 Created on Tue Jan 19 12:42:12 2021
4 
5 Author: Juanma
6 """
7 
8 from bio_seq import bio_seq
9 from DNA_structures import writeTextFile, readTextFile, read_FASTA
10 
11 fasta= read_FASTA("text.txt")
12 # print(fasta)
13 for f in fasta:
14 
15     seq=fasta[f]
16 
17 sarscov=bio_seq(seq)
18 sarscov.info_seq()
```

Terminal 1/A

TACATAGACAAGTAGATGTAGTTAACCTTAACTTCACATAGCAATTAACTCAGTGTAACATTAGGGGAGCTTGAAGAGGCCACCATTTACCGAGG
CCACGCGGATGAGATGCGAGTTACAGTGACCATGCTAGGGAGGCTTATATGGAAAGGCCCTAATGTGTAATAAATTTAGTGTATCCCCATGT
GATTTAAATAGCTTCTAAGGAATGACAAAAAAA
Tipo de cadena: DNA
Label: No label
Longitud de la cadena: 29903 nucleótidos

In [36]:

Vemos que el genoma del virus cuenta con apenas 30000 nucleótidos. No deja de resultar paradójico cómo una secuencia genética tan ínfima, y que de hecho cabe en apenas un par de páginas de los procesadores de texto, ha dado lugar a tan graves consecuencias.

Otros datos interesantes son los referidos a la proporción de nucleótidos y al contenido en GC

```

C:\Users\Juanna\Documents\Curso_Python\Bioinformática\ing protes.py
In [42]: runfile('C:/Users/Juanna/Documents/Curso_Python/Bioinformática/ing protes.py', wdir='C:/Users/Juanna/Documents/Curso_Python/Bioinformática')
Reloading modules: bio_seq, DNA_structures
GC content= 38%
{'A': 8954, 'C': 5492, 'T': 9594, 'G': 5863}
In [43]:

```

The screenshot shows a Jupyter Notebook interface with several tabs at the top: 'res.py', 'Ejecutando.py', 'text.txt', 'bio_func.py', 'bio_seq.py', and 'ing protes.py'. The 'ing protes.py' tab is active, displaying the following Python code:

```

1  # -*- coding: utf-8 -*-
2  #
3  # Created on Tue Jan 19 12:42:12 2021
4  #
5  # @author: Juanna
6  #
7  #
8  from bio_seq import bio_seq
9  from DNA_structures import writeTextFile, readTextFile, read_FASTA
10
11 fasta= read_FASTA("text.txt")
12 # print(fasta)
13 # for f in fasta:
14 #
15     seq=fasta[f]
16
17 sarscov=bio_seq(seq)
18 print(sarscov.gc_content())
19 print(sarscov.countar())
20
21

```

The notebook also includes a terminal window showing the execution of the script and its output. The output shows the GC content of the sequence as 38% and a dictionary of nucleotide counts: {'A': 8954, 'C': 5492, 'T': 9594, 'G': 5863}.

Es interesante observar que el porcentaje en GC es relativamente bajo. A pesar de que el contenido en GC puede parecer una característica que aporta poca información cada vez son mayores los indicios de que puede tener relación directa con el uso de codones y con la presión selectiva a la que está expuesto el genoma. De hecho, me parece especialmente interesante que este porcentaje se asemeje tanto al que tiene el genoma humano, en torno a un 40 por ciento, lo que me sugiere que determinadas proporciones en GC dan lugar a proporciones y usos de aminoácidos específicos que benefician la replicación del virus, ya que este debe ser procesado por la célula hospedadora.

Es más, determinados estudios indican que este porcentaje en GC está claramente optimizado y coincide con los porcentajes observados para los genes que más se expresan en el epitelio pulmonar. De esta forma, el virus habría conseguido potenciar su replicación en las células de este tejido. (Li y cols., 2020).

No obstante, tal vez lo más interesante pueda ser obtener la secuencia real del virus, que recordemos es ARN. Supongo que el hecho de que aparezca en “tipo ADN” se debe a la forma de obtención de la secuencia, pues por mis conocimientos entiendo que lo que se realiza es una RT-PCR para posteriormente proceder con la correspondiente secuenciación, pues las metodologías de Next Generation Sequencing aún no están lo suficientemente establecidas para llevar a cabo análisis de secuencia directamente del mRNA.

Por desgracia, a pesar de ser un virus con un material genético relativamente corto, casi 30000 nucleótidos sigue siendo una cantidad de información enorme, por lo que procederemos a tratar únicamente con la región correspondiente a nuestra proteína de interés, [19621 - 20658]. Ojo, hay que tener cuidado y restar un valor a cada posición, pues Python toma la posición inicial de una cadena como la posición 0.

The screenshot shows a Python development environment with two main panes. The left pane displays the source code for `ing protos.py`, which reads a FASTA file (`text.txt`) and prints its sequence. The right pane shows the terminal window with the output of running the script, including the loaded modules (`bio_seq, DNA_structures`) and the protein sequence itself.

```
# -*- coding: utf-8 -*-
"""
Created on Tue Jan 19 12:42:12 2021
@author: Juanma
"""

from bio_seq import bio_seq
from DNA_structures import writeTextFile, readTextFile, read_FASTA
fasta= read_FASTA("text.txt")
# print(fasta)
for f in fasta:
    seq=fasta[f]
seq=fasta[17]
sarscov=bio_seq(seq[19620:20657])
print("cadena real de la proteína: ")
cadena="HHHHHHSSGVLDGTENLYFQSNMISLENAVFNVNKGHDGQQGEVPVSIINNTVYTKV
print(cadena)
print()
cadena_new=""
print("cadena a partir de la información genética: ")
for a in sarscov.traducir():
    cadena_new+=a
print(cadena_new)
```

```
In [66]: runfile('C:/Users/Juanma/Documents/Curso_Python/Bioinformática/ing protos.py', wdir='C:/Users/Juanma/Documents/Curso_Python/Bioinformática')
Re-loaded modules: bio_seq, DNA_structures
cadena real de la proteína:
HHHHHHSSGVLDGTENLYFQSNMISLENAVFNVNKGHDGQQGEVPVSIINNTVYTKV
DGDV6DVELFNKTTLPVNVAELWAKRNKPVEVKILNNLGVDIAANT
VIIDYKRDAHIST7GVCMSMTDIACKPTETICAPLTVFDFGRDVQDVLFRNARINGVLITEGSVKGQLQPSPVGPQKASLNGVTLIGEAVKTQFYKKVVDGVVQQLP
ETYFTQSRLNQEFKPSQMEIDFLELANDIFIERYKLGEYAFEHIVYGFDSHSQQLGLHLILGAKRKFESPELEDFIPMDSTWKVYFITDAQTGSCKVCVSVID
LLDPDVEIKS0L5WSVSKVKTIDTYTEISFHLCKDGHVETYPKLQ

cadena a partir de la información genética:
SLENANFVNKGHDGQQGEVPVSIINNTVYTKV
DGDV6DVELFNKTTLPVNVAELWAKRNKPVEVKILNNLGVDIAANTVIMDYKRDAPAHIST7GVCMSMTD
IACKPTETICAPLTVFDFGRDVQDVLFRNARINGVLITEGSVKGQLQPSPVGPQKASLNGVTLIGEAVKTQFYKKVVDGVVQQLPETYFTQSRLNQEFKPSQMEIDFL
ELANDIFIERYKLGEYAFEHIVYGFDSHSQQLGLHLILGAKRKFESPELEDFIPMDSTWKVYFITDAQTGSCKVCVSVIDLLDDFVEIKS0L5WSVSKVKT
IDTYTEISFHLCKDGHVETYPKLQ

In [67]:
```

Se observa claramente que la proteína sufre ciertos cambios respecto a su secuencia, pero esto no se debe a una modificación o maduración, sino a que en los análisis que se realizan para determinar la estructura de la proteína los datos de los extremos suelen ser de menor calidad y no se tienen en cuenta para la creación del modelo.

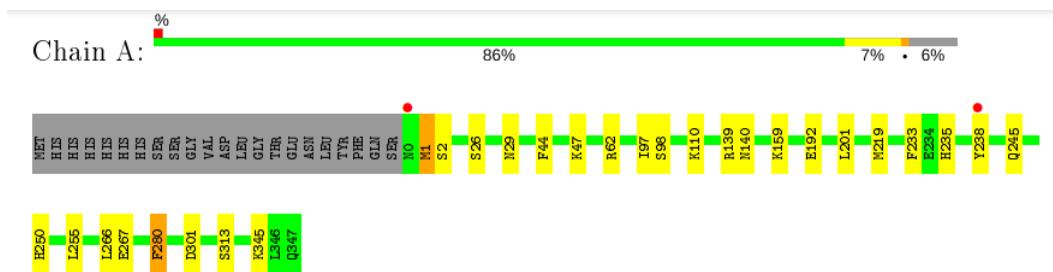


Figura 1: Imagen obtenida del documento correspondiente a la estructura 6WXC en PDB. Se indican en gris los residuos que no se representan en el modelo.

Con este pequeño análisis de la genética del virus hemos demostrado que los datos genéticos y estructurales tienen coherencia, de forma que podemos confiar más si cabe en los datos estructurales de las secuencias de PDB. Por supuesto, no es que yo desconfíe de los datos aportados en PDB, pero considero que una de las cualidades que debe tener todo investigador es la capacidad crítica de análisis de la información, por lo que nunca está de más entender de qué manera se obtienen los datos que se nos presentan para poder darles un uso adecuado. Con este punto de partida, procedemos a analizar los elementos estructurales de la proteína en cuestión.

Antes de proceder a analizar la estructura y conformación de la proteína cabe destacar que estamos trabajando sobre virus ARN, lo que tiene importantes repercusiones. En virus ARN no se suele hablar de especie, sino de cuasiespecie. Esto se debe a que en su proceso de replicación hay un gran cantidad de errores, lo que determina tasas de mutación mucho mayores a las observadas en seres vivos y virus tipo ADN, hasta cien mil veces mayores. Esto provoca que simultáneamente podamos encontrar diferencias bastantes significativas entre los genomas de una población de virus, y esto es lo que va a ocurrir con el SARS-COV2. La secuencia genética no se corresponde a una

variante predominante dentro de la población, sino a la secuencia consenso que se establece a partir de la estimación de todas las cepas y variantes circulantes.

Esto habrá que tenerlo en cuenta para los análisis estructurales, donde sí se posee una secuencia aminoacídica “real” obtenida de una de las variantes circulantes. La estructura observada será en sentido estricto la que utiliza el virus, es decir, una representación muy fiel, pero que habrá que ir revisando para tener en cuenta todas las modificaciones que se vayan produciendo. Hasta el momento no ha habido un gran número de mutaciones que se hayan expandido para la proteína NSP15, pero si tuviéramos que valorar otras como es el caso de la spike deberíamos valorar que recientemente se ha reportado un claro aumento en el acervo poblacional de determinadas mutaciones muy relevantes para el proceso de infección y replicación del virus.

2. Nsp15: Non Structural Protein 15

El nombre de esta proteína es el acrónimo de “Non Structural Protein 15” del SARS-COV2. Las proteínas no estructurales, como es de esperar, no forman parte de la estructura en sí misma del SARS-COV2 (estas proteínas serían la spike, la proteína de la membrana, la proteína de la envoltura y la de la nucleocápsida).

De las 27 proteínas identificadas, 16 son no estructurales, y han sido identificadas algunas proteínas accesorias que actúan frente al sistema inmune del hospedador. Centrándonos en las no estructurales sus funciones conocidas son varias y en general relacionadas con el procesamiento de las partículas de virus, si bien para algunas de ellas aún no se ha descrito una función aparente.

Específicamente, la NSP15, nuestra proteína de interés, se trata de una endonucleasa, es decir, su función es la de realizar cortes en el interior de secuencias génicas. No obstante, el conocimiento que poseemos actualmente sobre esta proteína es aún bastante limitado, lo que unido a sus peculiares características la hacen aún más enigmática.

Se trata de una riboendonucleasa específica de uridilato-RNA (del tipo NendoU al tratarse de un nidovirus) que posee un dominio C-terminal catalítico típico de la familia EndoU. Este tipo de proteínas EndoU se encuentran en todos los dominios de seres vivos (y por lo que parece, también en virus) y suelen tener funciones destinadas al procesamiento de distintos tipos de ARN. Además, actúan sobre ARN de cadena tanto simple como doble, y están presentes en diversos grupos de virus que infectan a vertebrados.

Más concretamente producen un corte que crea una estructura fosfodiéster 2'-3' ciclada dejando un 5' hidroxiterminal, y aunque inicialmente se pensó que actuaban de forma directa en la replicación viral posteriormente se ha podido demostrar que coronavirus deficientes en esta proteína eran capaces de llevar a cabo su ciclo de replicación. Es por ello que se ha propuesto un mecanismo de acción de la proteína frente al sistema inmune, en el que esta proteína destruye el ARN viral sobrante para que no sea detectado por las defensas del organismo infectado. No obstante, esta proteína es esencial para el correcto proceso infectivo, pues le permite al virus llevar a cabo una estrategia de “evasión” del sistema inmune (Kim, Jedrzejczak, y cols., 2020).

La entrada del virus en la célula hospedadora se lleva a cabo a través de la liberación de +gRNA (RNA genómico de sentido positivo) en el citoplasma. Desde aquí, este puede sufrir traducción, transcripción del -gRNA complementario, así como replicación y retrotranscripción. En este proceso la Nsp15 parece tener una función clave, pues permitiría una rápida degradación de las moléculas de RNA, con especial afinidad por las colas 3' poliU, que serían fácilmente detectadas por el sistema inmune del hospedador, desencadenando una respuesta inmune contraria al virus.

Respecto a la similitud en la secuencia aminoacídica, se ha demostrado que esta es amplia dentro del grupo de coronavirus (tales como el SARS-CoV1 y el MERS) mientras que las diferencias aumentan en gran medida cuando tenemos en cuenta otros virus relacionados.

Para comprobar esta información podemos utilizar la herramienta BLAST, tomando como secuencia de entrada la que hemos tomado anteriormente en formato FASTA.

Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>

Standard Protein BLAST

[blast](#) [blastp](#) [blastx](#) [tblastn](#) [tblastx](#)

BLASTP programs search protein databases using a protein query. [more...](#)

[Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#)

>YP_009725310.1 endoRNase [Severe acute respiratory syndrome coronavirus]

From _____ To _____

New columns added to the Description Table
Click 'Select Columns' or 'Manage Columns'. [New](#)

Or, upload file [Seleccionar archivo](#) Ningún archivo seleccionado

Job Title [YP_009725310.1 endoRNase \[Severe acute respiratory...](#)

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Lo primero que nos encontramos al ver las secuencias con mayor porcentaje de similaridad es que hay un gran número de secuencias que corresponden al SARS-COV2. Esto es debido a la inmensa cantidad de análisis que se están realizando sobre este virus, y además nos indica que esta proteína sufre muy pocas modificaciones y que en general, podemos considerar que nuestra secuencia consenso es adecuada.

Descriptions	Graphic Summary	Alignments	Taxonomy	Download	Select columns	Show	100	?	
Sequences producing significant alignments									
<input checked="" type="checkbox"/> select all 100 sequences selected	GenPept	Graphics	Distance tree of results	Multiple alignment	MSA Viewer				
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	ORF1ab polyprotein [Severe acute respiratory syndrome coronavirus 2]	Severe acut...	717	717	100%	0.0	100.00%	6867	QNG43344.1
<input checked="" type="checkbox"/>	ORF1ab polyprotein [Severe acute respiratory syndrome coronavirus 2]	Severe acut...	717	717	100%	0.0	100.00%	6924	QMV29246.1
<input checked="" type="checkbox"/>	ORF1ab polyprotein [Severe acute respiratory syndrome coronavirus 2]	Severe acut...	717	717	100%	0.0	100.00%	7096	QUO99120.1
<input checked="" type="checkbox"/>	ORF1ab polyprotein [Severe acute respiratory syndrome coronavirus 2]	Severe acut...	717	717	100%	0.0	100.00%	7096	QQG71325.1
<input checked="" type="checkbox"/>	ORF1ab polyprotein [Severe acute respiratory syndrome coronavirus 2]	Severe acut...	716	716	100%	0.0	100.00%	7096	QPD98106.1
<input checked="" type="checkbox"/>	ORF1ab polyprotein [Severe acute respiratory syndrome coronavirus 2]	Severe acut...	716	716	100%	0.0	100.00%	7096	QDD87628.1
<input checked="" type="checkbox"/>	ORF1ab polyprotein [Severe acute respiratory syndrome coronavirus 2]	Severe acut...	716	716	100%	0.0	100.00%	7096	QPF50724.1
<input checked="" type="checkbox"/>	ORF1ab polyprotein [Severe acute respiratory syndrome coronavirus 2]	Severe acut...	716	716	100%	0.0	100.00%	7096	OPN04968.1
<input checked="" type="checkbox"/>	ORF1ab polyprotein [Severe acute respiratory syndrome coronavirus 2]	Severe acut...	716	716	100%	0.0	100.00%	7096	QNN2013.1
<input checked="" type="checkbox"/>	ORF1ab polyprotein [Severe acute respiratory syndrome coronavirus 2]	Severe acut...	716	716	100%	0.0	100.00%	7096	QNO31842.1
<input checked="" type="checkbox"/>	ORF1ab polyprotein [Severe acute respiratory syndrome coronavirus 2]	Severe acut...	716	716	100%	0.0	100.00%	7096	QQG76433.1
<input checked="" type="checkbox"/>	ORF1ab polyprotein [Severe acute respiratory syndrome coronavirus 2]	Severe acut...	716	716	100%	0.0	100.00%	7096	QOJ39534.1

Sin embargo, si seleccionamos la opción de que no nos presente las similitudes con otras secuencias de la misma “especie”, los resultados son aún más interesantes.

Descriptions		Graphic Summary	Alignments	Taxonomy										
Sequences producing significant alignments									Download	New	Select columns	Show	100	?
<input checked="" type="checkbox"/> select all 100 sequences selected				GenPept		Graphics		Distance tree of results		Multiple alignment		New MSA Viewer		
		Description		Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession			
<input checked="" type="checkbox"/>	ORF1ab [synthetic construct]	synthetic construct	716	716	100%	0.0	100.00%	7096	QIG55856_1					
<input checked="" type="checkbox"/>	orf1ab polyprotein [Pangolin coronavirus]	Pangolin corona...	704	704	100%	0.0	97.69%	7089	QIG55944_1					
<input checked="" type="checkbox"/>	orf1ab polyprotein [Bat coronavirus RaTG13]	Bat coronaviru...	701	701	100%	0.0	97.69%	7095	QHR63299_2					
<input checked="" type="checkbox"/>	non-structural polyprotein 1ab [Bat SARS-like coronavirus]	Bat SARS-like c...	657	657	100%	0.0	89.02%	7073	ATO98118_1					
<input checked="" type="checkbox"/>	non-structural polyprotein 1ab [Bat SARS-like coronavirus]	Bat SARS-like c...	657	657	100%	0.0	89.02%	7073	ATO98143_1					
<input checked="" type="checkbox"/>	orf1ab polyprotein [Bat coronavirus]	Bat coronavirus	657	657	100%	0.0	89.02%	7073	ARI44798_1					
<input checked="" type="checkbox"/>	non-structural polyprotein 1ab [Bat SARS-like coronavirus WIV1]	Bat SARS-like c...	657	657	100%	0.0	89.02%	7073	AGZ48830_1					
<input checked="" type="checkbox"/>	non-structural polyprotein 1ab [Bat SARS-like coronavirus]	Bat SARS-like c...	657	657	100%	0.0	89.02%	7073	ATO98216_1					
<input checked="" type="checkbox"/>	orf1ab polyprotein [Bat SARS-like coronavirus]	Bat SARS-like c...	657	657	100%	0.0	89.02%	7063	AID16715_1					
<input checked="" type="checkbox"/>	ORF1ab [Coronavirus BiRs-BetaCoV/YN2018B]	Coronavirus BiR...	657	657	100%	0.0	89.02%	7073	QDF43824_1					
<input checked="" type="checkbox"/>	non-structural polyprotein 1ab [Bat SARS-like coronavirus RsSHC014]	Bat SARS-like c...	657	657	100%	0.0	89.02%	7073	AGZ48805_1					
<input checked="" type="checkbox"/>	non-structural polyprotein 1ab [Bat SARS-like coronavirus]	Bat SARS-like c...	657	657	100%	0.0	89.02%	7073	ATO98106_1					

Las secuencias más similares y por tanto, más cercanas a nivel filogenético corresponden a coronavirus que infectan a pangolines y murciélagos. Esto es lo que se ha estado defendiendo desde hace meses por la comunidad científica, y con apenas una breve búsqueda hemos podido comprobar en qué se basan para proponer esta hipótesis los expertos que trabajan en este campo de investigación, muy alejados de las teorías conspiranoicas que aseguran sin ninguna base científica que el virus no ha surgido de forma natural. Según los datos recogidos en este somero análisis en BLAST podríamos afirmar que nuestro trabajo tiene cierta coherencia.

Otro factor interesante a tener en cuenta es que las mayores similitudes no se dan con las secuencias del MERS y SARS-COV1 que crearon ciertos problemas hace unos años, sino con las de coronavirus que infectan a otras especies. Los virus del MERS y SARS-COV1 no consiguieron prosperar como sí lo ha hecho el SARS-COV2, y la clave podría estar en las diferencias genéticas y proteicas que presenta respecto a sus predecesores (R. Kumar y cols., 2020). Por ello, el análisis de la estructura de la proteína lo haremos tomando como punto de referencia las proteínas homólogas de estos virus anteriores, pues a pesar de infectar a humanos no dieron lugar a una emergencia sanitaria como la actual, y podremos inferir los efectos de las mutaciones y diferencias estructurales que podamos observar.

Más concretamente y centrándonos en nuestra proteína de interés, los

cambios no son de una gran envergadura, pero se ha hipotetizado que serían suficientes para contribuir sustancialmente a la virulencia del SARS-COV2.



Figura 2: Superposición de los monómeros de la Nsp15 del SARS-COV2 (en verde azulado), el SARS-COV1 (rosa) y el MERS (gris). Tomada de Kim et al. (2020)

Sin embargo, para poder entender estos cambios primero debemos centrarnos en analizar la estructura de la proteína, entender sus dominios y regiones de interés y posteriormente, inferir la funcionalidad y características de esta secuencia aminoacídica.

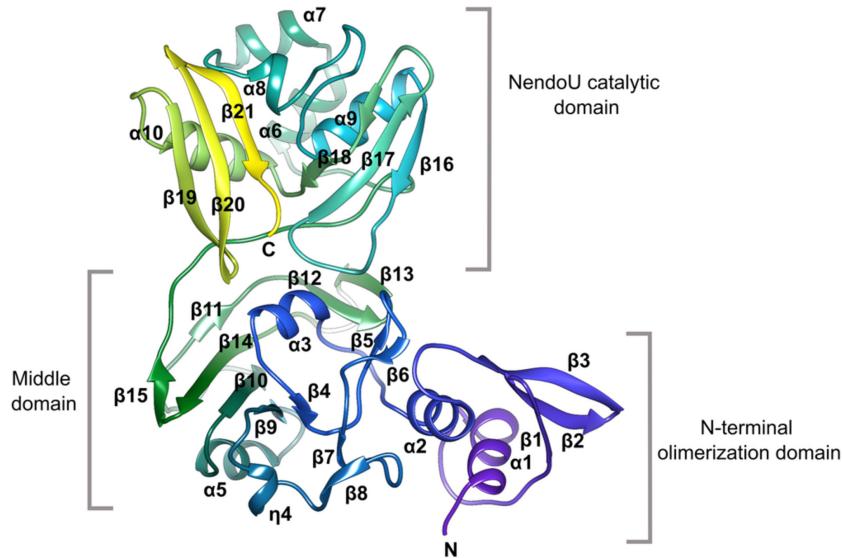


Figura 3: Estructura del monómero de la proteína Nsp15 con sus principales dominios y regiones estructurales

Como podemos observar, el monómero presenta 3 dominios principales:

- 1. Dominio NendoU donde reside la actividad catalítica (C-terminal)
- 2. Dominio intermedio que une los dominios N y C-terminales.
- 3. Dominio N-terminal responsable de la formación del hexámero

2.1. Dominio NendoU

En este dominio es donde reside el centro activo catalítico responsable de la actividad endonucleasa sobre ARN de doble cadena. Más concretamente, en una pequeña hendidura que se encuentra entre las β láminas de este dominio, donde destacan 6 residuos (His235, His250, Lys290, Thr341, Tyr343 y Ser294), con una posición y arquitectura conservada entre los 3 coronavirus mencionados a excepción de la Lys290, que varía ligeramente su disposición. (Pillon y cols., 2021)

His235, His250 y Lys290 son considerados como la triada catalítica, debido a que sus posiciones concuerdan con la que presenta la ribonucleasa A o pancreática, una ribonucleasa poco específica muy estudiada. Por su parte,

la Ser294 y la Tyr343 serían los responsables de controlar la especificidad, pues ocupan la posición de los aminoácidos responsables de esa acción en la ribonucleasa A. De hecho, se ha propuesto que la Ser294 sería la encargada de interactuar con el átomo de *O* del grupo carbonilo del uracilo a través del *N* de la cadena principal, mientras que el grupo hidroxilo interaccionaría con el átomo de *N*.

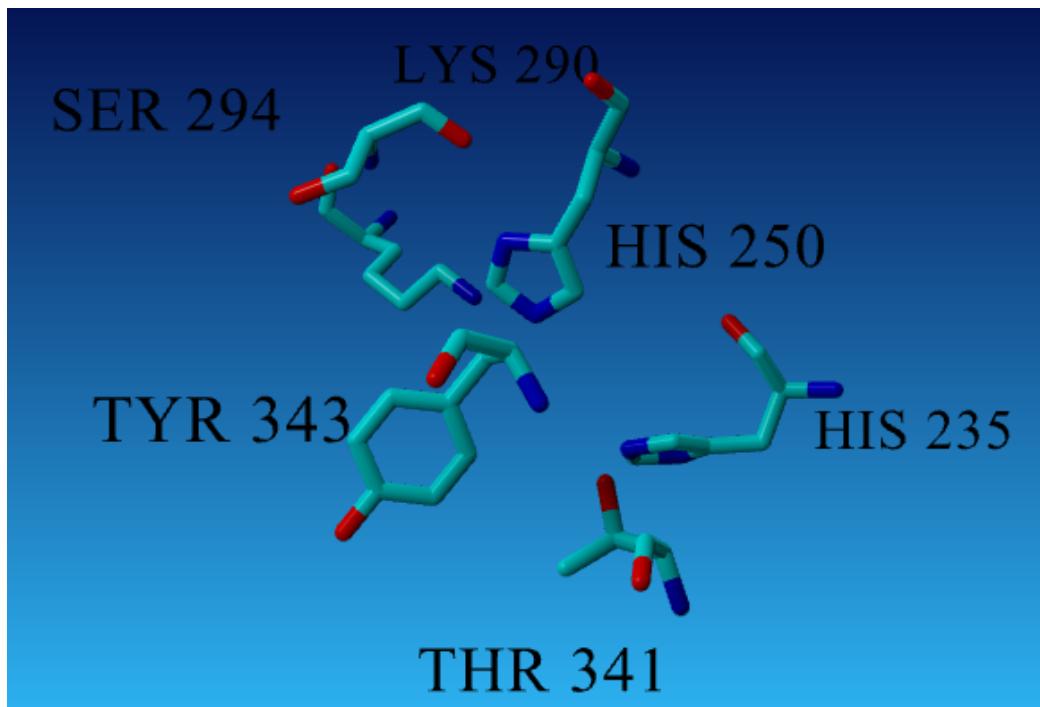


Figura 4: Imagen obtenida del software de visualización YASARA, donde se muestran los principales residuos que forman parte del centro activo con actividad endonucleasa. En la hendidura que se forma se podría introducir el Uracilo con el que se llevaría a cabo la excisión de la cadena ribonucleica.

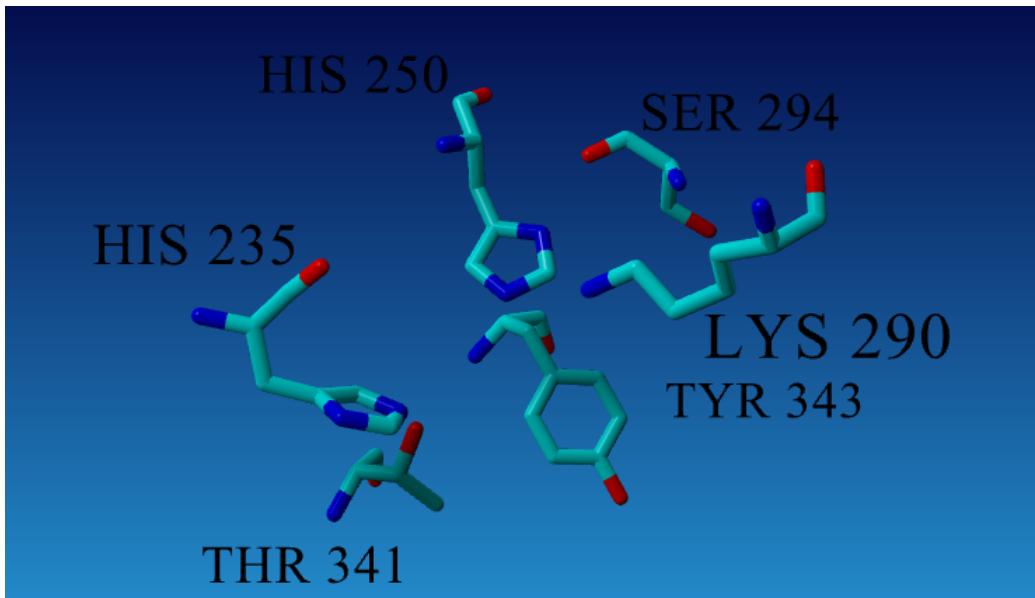


Figura 5: Imagen desde el ángulo opuesto, que nos permite entender la estructura tridimensional del centro activo.

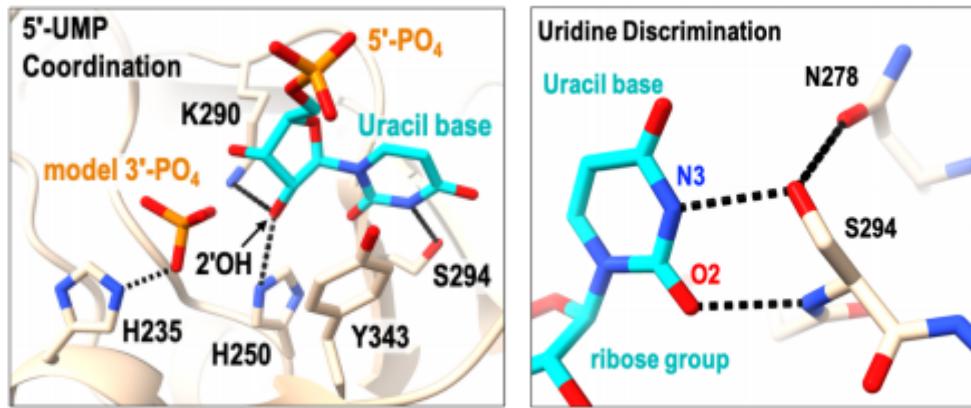


Figura 6: Interacción del uracilo con el centro activo de la proteína y su disposición dentro del mismo. Especialmente destacable el papel de la Ser294

2.2. Dominio intermedio

Se ha hipotetizado sobre la participación de este dominio en la formación del hexámero (forma activa de la proteína). Al parecer, contribuye a la

creación de una estructura cóncava que permitiría la introducción de ARN y proteínas y la interacción de la Nsp15 activa con estos, lo que podría repercutir en una mejora en la actividad catalítica. No obstante, la información es escasa en lo referente a la proteína en su forma hexamérica.

A pesar de que este pudiera parecer el dominio menos interesante de la Nsp15 es el que presenta un mayor grado de diferencias respecto a sus proteínas homólogas del SARS-COV1 y el MERS. Estos cambios podrían deberse a que no son regiones tan importantes para la función de la proteína y por tanto, no se encuentran tan conservadas. No obstante, ha crecido en importancia la hipótesis de que estas mutaciones podrían tener un papel esencial en el desarrollo de una estructura más eficiente del hexámero. El principal problema para corroborarla, sin embargo, es que no ha sido posible determinar las condiciones necesarias para la hexamerización, si bien existen diversas hipótesis.

2.3. Dominio de oligomerización

Debido a las similitudes que presenta la Nsp15 del SARS-COV2 con otras proteínas NendoU se cree que para que se encuentre en forma activa debe formar un oligómero compuesto de 6 subunidades. Para ello, el dominio de oligomerización es esencial, pues a través de este es donde se dan las principales interacciones para tal fin.

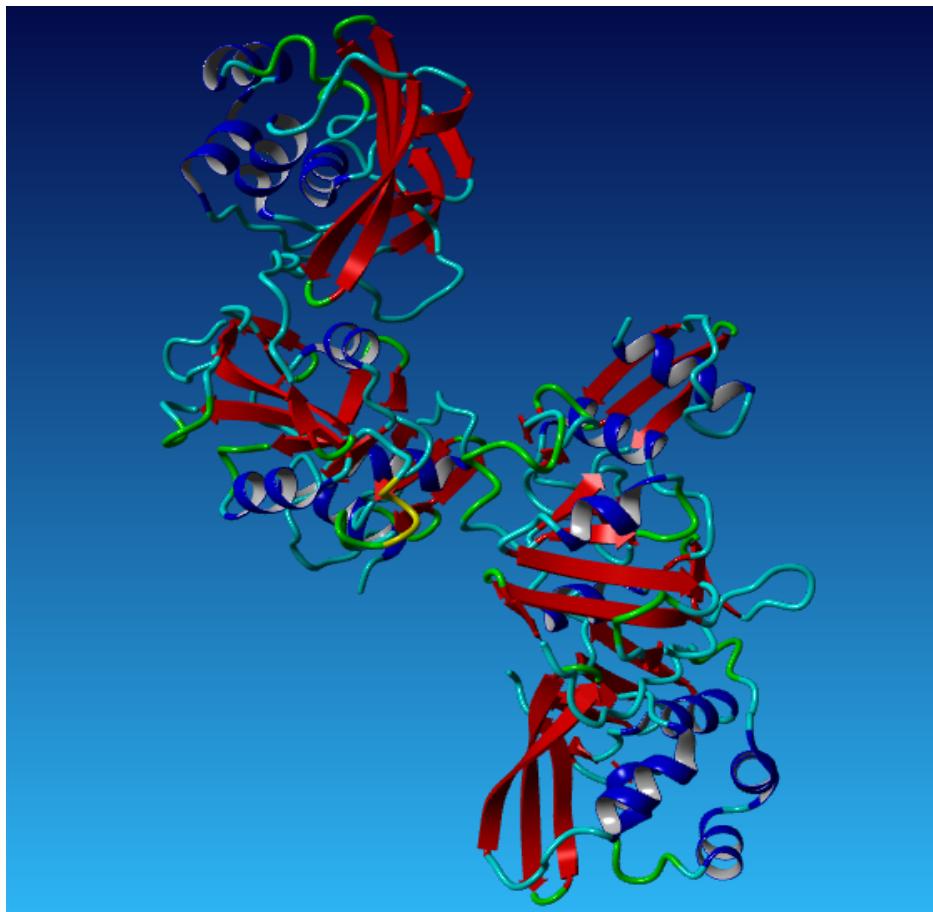


Figura 7: Imagen obtenida del software de visualización YASARA, donde se puede observar la interacción entre dos subunidades de la Nsp15 a través del dominio de oligomerización. Las subunidades se disponen en sentido opuesto

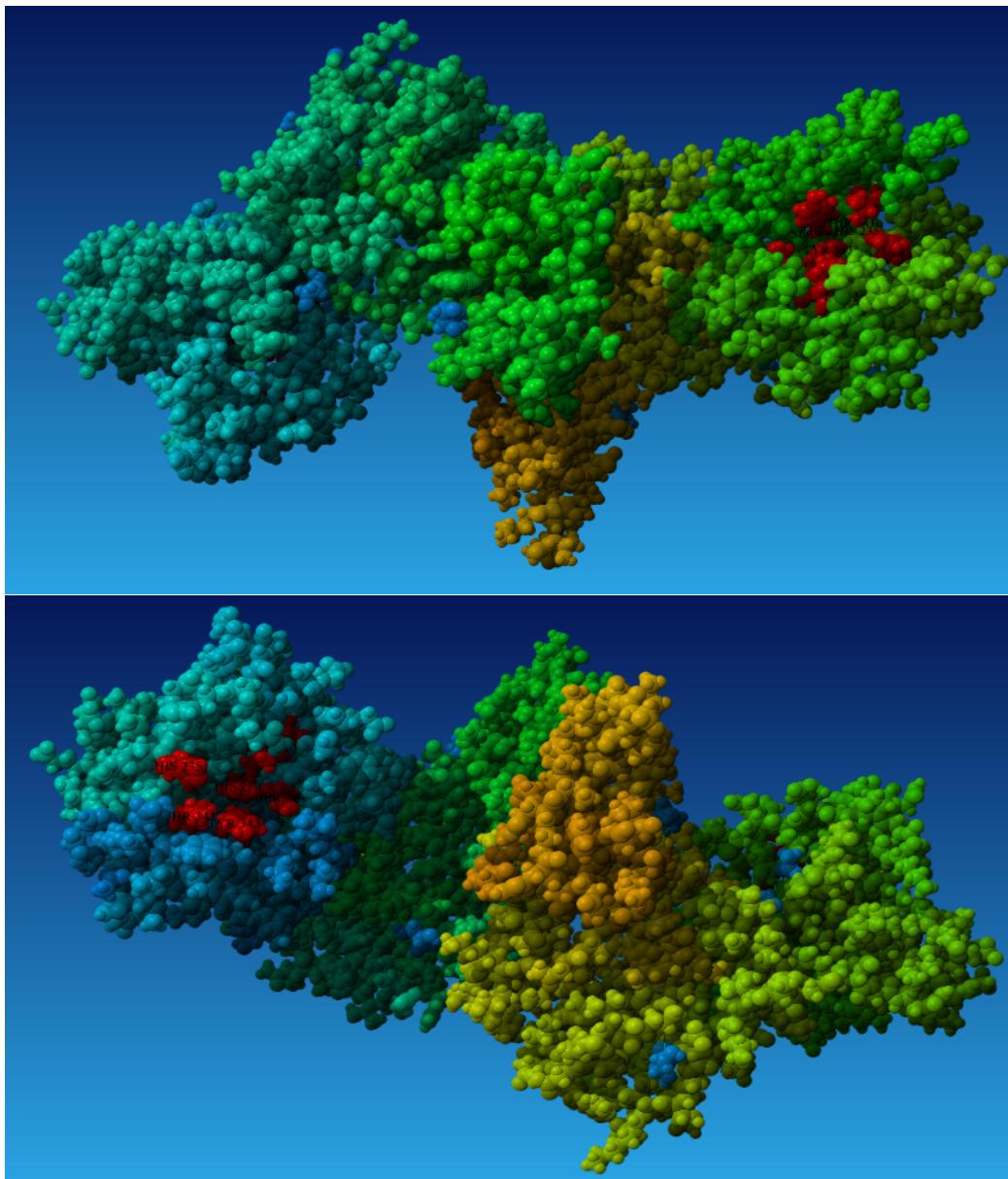


Figura 8: Imágenes obtenidas desde distintos ángulos donde aparecen marcados los residuos del centro activo (rojo), sirviendo de referencia para entender la disposición de los dominios catalíticos de cada subunidad, pudiendo observar la unión por las regiones de oligomerización. Además, se puede observar con gran facilidad la hendidura del centro activo.

Por otra parte, las predicciones estructurales parecen indicar que todos los dominios de cada subunidad interaccionan para permitir la formación del dímero de trímeros que constituye la forma activa de la Nsp15. Además, los indicios indican que todas y cada una de las subunidades interaccionan en mayor o menor medida con el resto, lo que crearía una estructura muy compacta.

Debido a que no se ha encontrado ninguna divergencia importante es las regiones encargadas de la actividad catalítica de la enzima los investigadores piensan que las principales modificaciones en la secuencia y estructura de la proteína han tenido como resultado una mejor disposición de las subunidades dentro del hexámero, lo que repercutiría en una acción más eficiente del complejo en el corte de cadenas ribonucleicas.

2.4. Hexámero

La forma hexamérica es la única en la que la Nsp15 presenta actividad. No obstante, y a pesar de que se desconocen las condiciones necesarias para la formación del mismo, se ha propuesto un modelo en el que la activación de la proteína se produce en base a un cambio alostérico en la misma dependiente de la oligomerización. Más específicamente, se ha observado que la oligomerización tendría repercusiones en la disposición del centro activo, restringiendo su movilidad y por tanto, el número de estados posibles.

Además, se ha observado que la unión de UMP modifica la arquitectura del centro activo, estableciendo aún más similitudes con la bien estudiada ribonucleasa A, si bien parece ser que daría lugar a una estructura con una capacidad catalítica relativamente lenta.

Esto tiene vital importancia. El dominio NendoU tiene una alta movilidad, similar a un balanceo. Se cree que esto permitiría una rápida unión a las moléculas de ARN, mientras que la reacción sería lenta. Esto facilitaría una retirada rápida de moléculas que podrían ser reconocidas por el sistema inmune del hospedador. No obstante, también hay indicios de que este balanceo pueda ser clave en la comunicación con el resto de subunidades que forman el hexámero, a través de cambios alostéricos. De hecho, la hipótesis más aceptada actualmente indica que la estructura en hexámero con disposición simétrica en un anillo de dímero de trímeros sería el punto de partida

para un mayor número de conformaciones, todas ellas asimétricas, que permitirían la unión a diferentes puntos de un mismo ARN.

De esta manera sería posible explicar que la conformación en hexámero sea la única activa para la proteína Nsp15. No obstante, aún quedan muchas cuestiones sin resolver. Entre ellas, destacan algunas referentes a si todos los centros activos realizan cortes o si simplemente pueden servir de anclaje, lo que permitiría un corte más específico en la longitud de las colas poliU, o incluso el ya mencionado problema sobre cómo se forma este hexámero y en qué condiciones lo hace. A pesar de que se han descrito numerosos ligandos y moléculas capaces de unirse a la proteína, su acción sigue sin estar clara, lo que le añade aún más misterio a esta estructura.

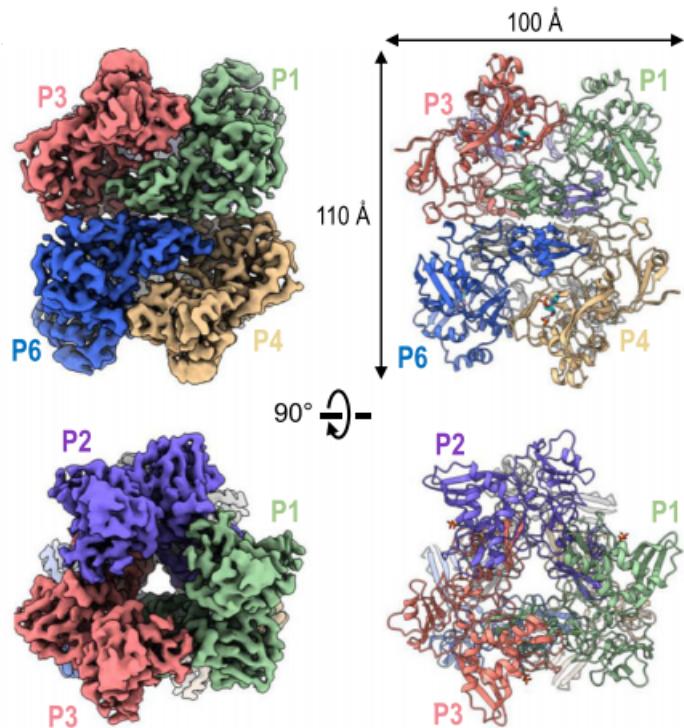


Figura 9: Estructura en anillo de dímero de trímeros descrita para la Nsp15. Obtenida de Pillon et al. (2021)

3. Opciones terapéuticas

Debido a la situación de crisis sanitaria, gran parte de la comunidad científica se ha volcado en el desarrollo de medicamentos y vacunas para hacer frente al SARS-COV2.

En torno a la casuística de nuestra proteína en estudio, el desarrollo de una vacuna no parece adecuado, pues no se trata de una proteína estructural, por lo que esta opción quedaría descartada. Más aún si comparamos con otras proteínas presentes en el SARS-COV2 como puede ser la spike, protagonista en las diversas vacunas que están superando los ensayos, consiguiendo su aprobación y llegando al mercado.

Sin embargo, el desarrollo de medicamentos que tengan como target la Nsp15 sí que se ha valorado y existen diversos estudios al respecto. En general, las ribonucleasas son proteínas ampliamente presentes en las células humanas y cumplen diversas funciones, por lo que no será posible la utilización de un inhibidor general de endoribonucleasas. En su lugar, se debe buscar una mayor especificidad, con el objetivo de no alterar el resto de procesos de las células humanas, pues esto podría suponer una fuente de toxicidad e inhabilitaría su utilización terapéutica.

No obstante, como hemos visto previamente, el centro activo de la enzima se asemeja en gran medida a la ribonucleasa A, que precisamente es una de estas riboendonucleasas de acción general. A pesar de esto he logrado encontrar algún estudio en el que tratan de utilizar inhibidores de ribonucleasa A *in vitro* para probar su eficacia frente a la Nsp15 del SARS-COV1 y obtienen resultados (Ortiz-Alcantara y cols., 2010). Por desgracia, creo que no se ha tenido en cuenta la posible toxicidad de dichos compuestos y no se han realizado más ensayos al respecto, por lo que dudo de su posible aplicación. Por ello, considero que la especificidad y mejores resultados vendrán dados por la unión que determina la Ser294 que hemos mencionado previamente en este informe. El objetivo será obtener un medicamento que actúe como competidor de las cadenas de RNA con alto contenido en uracilo.

De esta manera, sería posible en un paciente con la infección activa disminuir la capacidad del virus para “esconderse” del sistema inmune, dejando libre gran cantidad de ARN exógeno que se podría reconocer como extraño

y desencadenando una respuesta inmunológica adecuada para contenerlo y evitar el desarrollo de síntomatología grave.

3.1. Saikosaponinas

Las saikosaponinas son derivados del oleanano, un triterpenoide natural, que han sido ampliamente utilizadas en la medicina tradicional oriental. Se obtenían principalmente de plantas, siendo ingeridas a través de infusiones y su acción terapéutica frente a virus ha sido constatada en multitud de ocasiones.

Para hacer frente a la COVID-19 se han realizado estudios in silico, tratando de determinar si algunas de estas saikosaponinas podrían servir como inhibidoras de determinadas proteínas del SARS-COV2, es especial de nuestra Nsp15. Especialmente destacable es el caso de saikosaponina V, con la cual se observó que podía unirse al centro activo de la Nsp15, bloqueando por consiguiente la unión de cadenas de ARN, y complicando al virus su acción de evasión del sistema inmune. Adicionalmente, cabe destacar que se observó que esta saikosaponina junto con la de tipo U podrían unirse también a la proteína spike e impedir su función (Sinha y cols., 2020). De ser exitosa esta estrategia su potencial sería enorme. Por una lado inhibiría la entrada del virus a la célula, y por otra permitiría el reconocimiento del ARN viral desencadenando la respuesta inmune pertinente.

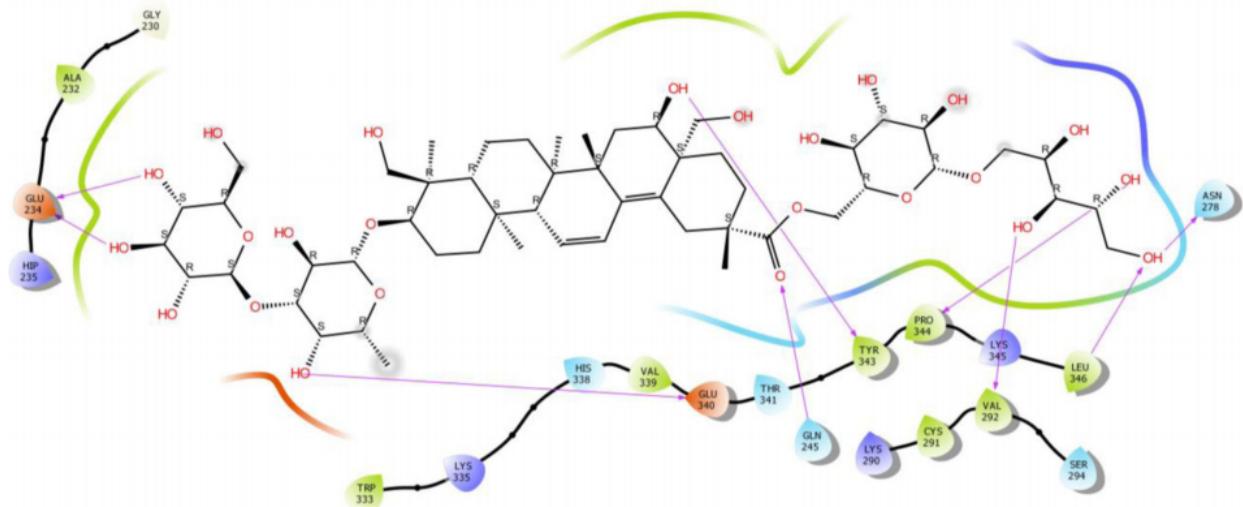


Figura 10: Representación de las interacciones planteadas entre la saikosaponina V y la Nsp15. Sinha y cols. (2020)

No obstante, cabe destacar que se trata de un estudio *in silico* y que debería replicarse en condiciones de laboratorio para confirmar su efectividad.

Siguiendo esta línea de investigación se han propuesto diversos compuestos químicos derivados de plantas para hacer frente al SARS-CoV2. De hecho, en este ámbito la bioinformática está desempeñando un papel crucial, pues está permitiendo obtener un gran número de resultados de posibles interacciones entre moléculas a través de software de auto-docking (Savale y cols., 2021) (Batool, Bibi, Amin, y Kamal, 2021), lo que nos permite hacer un screening inicial de posibles candidatos. Gracias a esta estrategia se han propuesto diversos compuestos útiles: demetoxicurcumina, quercetina y miricetina entre muchos otros. (Sharma, Goyal, Yadav, Kumar, y Gupta, 2020) (S. Kumar y cols., 2020) (Umar, Siraj, Ajayi, Jimoh, y Chukwuemeka, 2021)

No obstante, no he encontrado información referente a ensayos clínicos con estos compuestos, por lo que entiendo que se están tomando en cuenta otras vías prioritarias. Aún así, es importante continuar con estas investigaciones, pues estamos observando nuevas mutaciones en el SARS-CoV2 que amenazan a la efectividad de las vacunas, y tampoco sabemos cuánto tiempo va a permanecer el virus entre nosotros.

3.2. Tipiracil

El tipiracil es un análogo sintético del uracilo, aprobado desde hace ya un tiempo por la FDA para el tratamiento del cáncer colorrectal que presenta una potencial aplicación frente a la Nsp15 (Kim, Wower, y cols., 2020).

La estrategia es simple, el tipiracil podría actuar como inhibir competitivo del uracilo y unirse al centro activo de la proteína, lo que impediría el ingreso de las cadenas de RNA y por tanto, dificultaría la evasión del sistema inmune al SARS-COV2. Con técnicas de molecular docking los resultados eran esperanzadores, y se provó su funcionalidad in vitro, obteniendo una ligera disminución en la replicación del virus. Puede no parecer un gran comienzo, pero la realidad es que ahora mismo no existe ningún medicamento con acción terapeútica real contra la COVID-19 cuando esta se encuentra activa en el paciente, por lo que seguir investigando con este compuesto podría llevarnos a, al menos, tener una manera de mitigar la enfermedad en determinados pacientes.

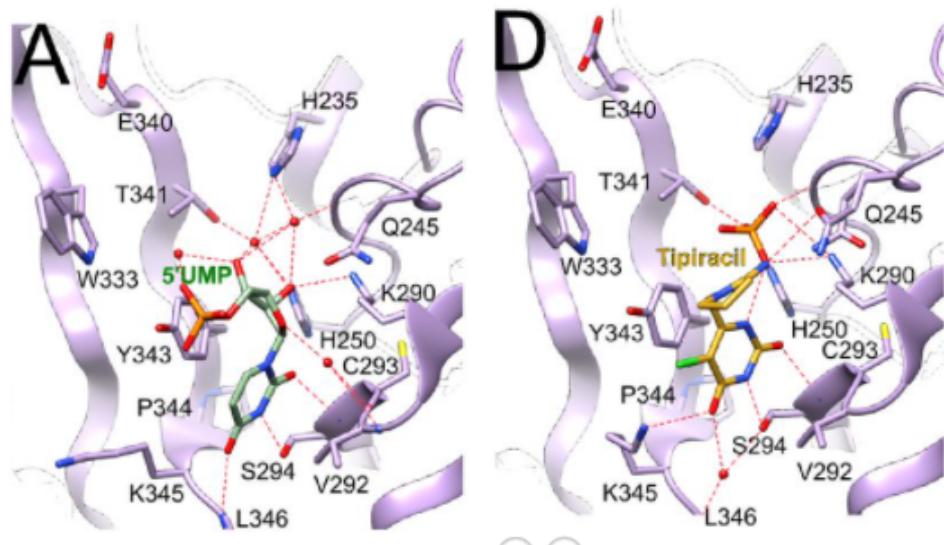


Figura 11: Comparación de la unión del UMP y el tipiracil

```

tipo:='No se detecta el tipo de archivo';
if (copy(memol.lines[0],0,6)='HEADER') then tipo:='PDB'
else if (copy(memol.lines[0],0,5)='LOCUS') then tipo:='GenBank'
else if (copy(memol.lines[1],0,2)='AC') then tipo:='Uniprot'
else if (copy(memol.lines[1],0,2)='XX') tipo:='Emsemble';
edit1.caption:= tipo;

```

Figura 12: Unión del tipiracil a los centro activos de la proteína en su forma hexamérica

Al parecer el Tipiracil es capaz de estabilizar los residuos del centro activo, lo que crea una reordenación específica de la proteína. Además, se ha observado una importante acción de la Ser294, tal y como argumentaban otros artículos, en la especificidad de acción de la proteína. Además, parece que la unión a cada centro activo funciona de forma independiente, por lo que sigue sin estar clara la razón por la cuál es necesaria la estructura en hexámero para la funcionalidad de la proteína. En definitiva, el tipiracil puede suponer una importante opción de cara a combatir la infección por el covid.

3.3. Ciclenosida

Este compuesto químico es un glucocorticoide inhalado. Este tipo de moléculas se plantearon para hacer frente a la COVID-19 por una doble vía. Por un lado para hacer frente a la replicación del coronavirus, y por otra para disminuir la inflamación del paciente, provocada por la llamada "tormenta de citoquinas". Dentro de los esteroides que se han estudiado, la ciclenosida mostró una baja toxicidad y una capacidad notoria en la disminución de la replicación viral, y además lo hacía de forma específica contra los coronavirus y no contra otros como la gripe, lo que aporta una mayor seguridad sobre su especificidad. (Matsuyama y cols., 2020)

Se observó también que una determinada mutación en la Nsp15 creó resistencia al tratamiento con ciclenosida in vitro, lo que permitió proponer un mecanismo de acción donde la ciclenosida actúa sobre esta proteína, ya sea directa o indirectamente.

Estudios posteriores in silico han tratado de descifrar la acción de la ciclenosida, observándose una posible unión al sitio catalítico, en contra a otras

hipótesis que indicaban una posible acción sobre el dominio de dimerización en la formación del hexámero (Kimura y cols., 2020).

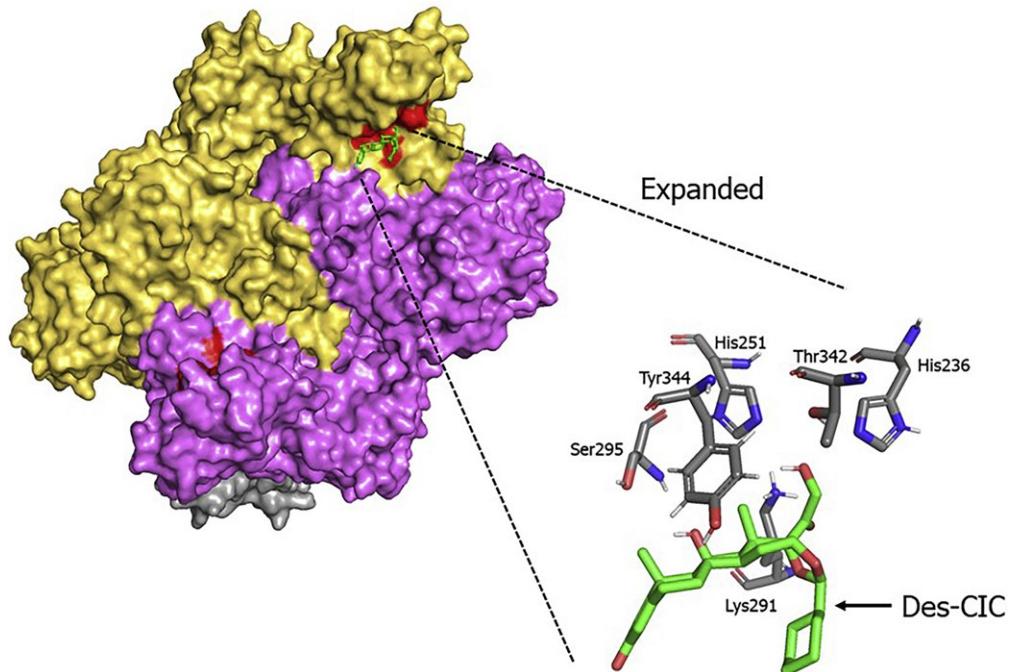


Figura 13: Unión de la ciclenosida al centro activo de la Nsp15

A pesar de todo, el mecanismo sigue sin estar claro, y el hecho de que se inhibía relativamente la replicación viral sigue sin dejar claro si esto se produce directamente por acción de la ciclenosida, o de forma indirecta por exposición al sistema inmune. No obstante, hay voces que apuntan a que la Nsp15 podría actuar por una doble vía, modificando RNA para mejorar su procesamiento y a la vez destruyendo el exceso para evitar una respuesta inmunológica.

4. Actividades

4.1. Bases de datos secuenciales y estructurales

4.1.1. Diagrama de flujo

El primer ejercicio consiste en la realización de un diagrama de flujo que nos permita distinguir entre distintos tipos de archivos, donde destacamos 4 de los más importantes utilizados por la comunidad científica en lo referente a la secuencia y estructura de proteínas: EMBL, Uniprot, GenBank y PDB. Para ello debemos estudiar la estructura y formato que siguen este tipo de archivos. Las diferencias entre ellos serán las que nos permitirán discernir entre cada tipo de archivo, y crear unos comandos lógicos para que nuestra aplicación pueda responder adecuadamente al tipo de archivo que se le indica.

Si bien podríamos hacer una gran revisión de la estructura de cada uno de estos cuatro tipos de archivos, nos centraremos únicamente en los que nos permiten su identificación y la determinación de la secuencia.

Respecto a la elección del tipo de archivo, los parámetros a tener en cuenta son los siguientes:

- PDB. Su primera línea comienza por “HEADER”
- GenBank. Su primera línea comienza por “LOCUS”
- Uniprot. Su primera línea comienza por “ID”, y la segunda por “AC”.
- Ensemble. Su primera línea comienza por “ID”, y la segunda por “XX”.

Lo que haremos será proceder a crear una estructura lógica de tipo condicional en Pascal, de acuerdo a la siguiente imagen.

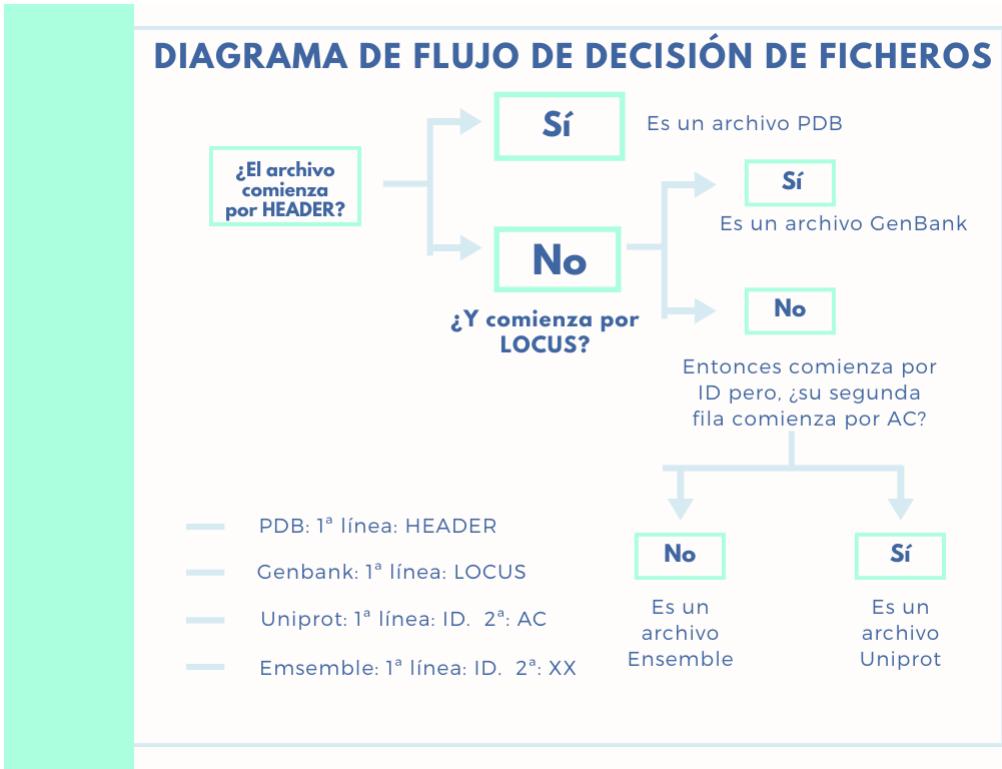


Figura 14: Diagrama de flujo con el que hemos ideado el programa que determina el tipo de fichero

Para ello compararemos las líneas del memo que hemos creado con las correspondientes a cada tipo de archivo. Adicionalmente, si no se detecta ninguno de los tipos de archivo el tipo de archivo se queda por defecto como "No detectado", lo que nos permite controlar errores y dirigir al usuario.

```

tipo:='No se detecta el tipo de archivo';
if (copy(memol.lines[0],0,6)='HEADER') then tipo:='PDB'
else if (copy(memol.lines[0],0,5)='LOCUS') then tipo:='GenBank'
else if (copy(memol.lines[1],0,2)='AC') then tipo:='Uniprot'
else if (copy(memol.lines[1],0,2)='XX') tipo:='Emsemble';
edit1.caption:= tipo;

```

Figura 15: Estructura condicional utilizada

Paralelamente, se puede obtener la secuencia de la proteína para tres

tipos de archivo: PDB, GenBank y Uniprot. Para ello simplemente habrá que utilizar el seleccionador de archivo anterior y delimitar, a partir de la estructura de cada tipo de archivo, obtener la secuencia aminoacídica.

- PDB. En este caso ya podemos obtener la secuencia gracias a que en Biotools ya habíamos desarrollado esta opción.
- GenBank. Buscaremos la línea que comienza por “/translation=”, que determina el comienzo de la secuencia, y la línea que comienza por “ORIGIN”. La secuencia comienza en la primera línea y termina en la anterior a la de “ORIGIN”.
- Uniprot. En este caso la secuencia está al final del fichero, comenzando en la línea posterior a la que comienza por “SQ”. Por ello, rastrearemos el memo hasta encontrar esa línea y copiaremos desde la línea siguiente hasta el final del mismo.

4.1.2. Cargar PDB

Para trabajar adecuadamente con nuestra proteína, necesitaremos leer correctamente su fichero correspondiente. Dentro de los más comunes utilizados en el estudio de proteínas, los archivos de extensión “PDB” son probablemente los más utilizados. Estos archivos tienen la características de ser estructurados y contener la información de la posición de cada átomo y la secuencia de la proteína, así como otros datos relacionados con la calidad de la información apartada y cómo se ha obtenido la misma.

Para aprovechar esta información hemos creado una función denominada “CargarPDB”, la cual nos creará un objeto de tipo “TPBD”, que hemos denominado así porque contendrá toda la información relacionada con la estructura y posición de los átomos, los residuos y la subunidad de la proteína.

Es muy importante conocer la estructura de este objeto, pues va a ser la pieza angular de todas las acciones que realicemos en el resto de aplicaciones y además, será clave para cualquier representación que hagamos ya que alberga toda la información estructural y posicional.

En líneas generales, está estructurada en tres espacios diferenciados: datos referentes a los átomos, datos referentes a los residuos, y datos referidos

a las subunidades, de menor a mayor nivel de complejidad.

Para ello, hemos tomado la información de las cadenas que comienzan por “ATOM” para la información relativa a los átomos (obteniendo sus coordenadas, código ID, residuo, subunidad, e incluso temperatura). Esto es lo que denominamos TAtomPDB, y que nos permite acceder a las características de cada átomo introducido.

Tras esto, para los residuos hemos tomado todos los átomos con ID “N” cabecera de residuo, es decir, de la cadena principal, los de la cadena lateral tienen otra nomenclatura diferente, obteniendo los ID del residuo de 3 y 1 letra, el nº de residuo y la subunidad a la que pertenece, así como hemos agrupado todos los átomos de ese residuo e incluso sus ángulos ϕ y ψ). Esto es lo que denominados TResiduoPDB

Por último, hemos creado una estructura que contiene las subunidades a partir de los residuos cabecera de cada subunidad, y que tiene integrada la información necesaria para referenciar a los átomos y residuos de las estructuras anteriores. Esto es lo que denominados TsubunidadPDB.

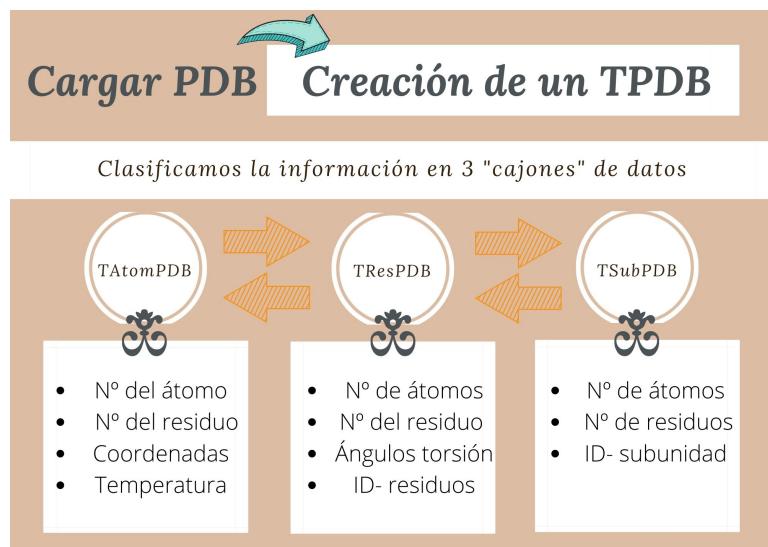


Figura 16: Utilidad y estructura de la función “CargarPDB”

4.2. Visualización de proteínas

4.2.1. Ficheros de estructuras

Esta actividad ha sido llevada a cabo de forma paralela a la revisión bibliográfica, utilizando los ficheros “6VWW” y “6WXC” pues el último archivo “6X41” no se encuentra aún disponible en la base de datos del Protein Data Bank. En el apartado Nsp15: Non Structural Protein 15 hemos incluido las imágenes y el análisis de las mismas para complementar la información obtenida en la búsqueda bibliográfica.

4.2.2. WritePDB

Lo que trataremos de hacer en esta ocasión, a la inversa que en la función CargarPDB, es tomar los datos de un TPDB, y con ello recrear” las líneas correspondientes a los átomos, con su misma estructura de forma que puedan ser leídos por un programa de visualización de imágenes. Para ello es clave tener claro cuál es la arquitectura de una línea ATOM de un archivo PDB y así poder respetar los caracteres y espacios de cada parámetro.

Para ello, de igual manera que hemos trabajado en la actividad para determinar los tipos de archivo, comprobamos la estructura de los ficheros PDB. Además, el ejercicio nos pide exclusivamente seleccionar los C α de la proteína.

Tras crear la función WritePDB y un par accesorias para crear un encolumnamiento adecuado procedemos a crear una aplicación que nos permite obtener en un memo la secuencia de líneas requerida con los correspondientes C α de la proteína de interés.

Cargar proteína		WritePDB	
<input type="file" value="mantis\ingeniería de proteínas\6www.pdb"/>		<input type="button" value="WritePDB"/>	
<pre> ANISOU 5549 O TYR B 343 4334 3793 7855 -94 -5 ^ ATOM 5550 CB TYR B 343 -48.684 28.540 29.656 1.00 ANISOU 5550 CB TYR B 343 5398 3946 9140 -191 -9 ATOM 5551 CG TYR B 343 -47.482 28.718 28.754 1.00 ANISOU 5551 CG TYR B 343 5093 3810 8938 -511 -8 ATOM 5552 CD1 TYR B 343 -46.217 28.935 29.286 1.00 ANISOU 5552 CD1 TYR B 343 5106 4044 9160 -682 -9 ATOM 5553 CD2 TYR B 343 -47.615 28.681 27.371 1.00 ANISOU 5553 CD2 TYR B 343 5022 3747 8748 -686 -7 ATOM 5554 CE1 TYR B 343 -45.118 29.104 28.469 1.00 ANISOU 5554 CE1 TYR B 343 5000 4214 9222 -1012 -8 ATOM 5555 CE2 TYR B 343 -46.520 28.849 26.545 1.00 ANISOU 5555 CE2 TYR B 343 5221 4215 8999 -1052 -7 ATOM 5556 CZ TYR B 343 -45.273 29.062 27.100 1.00 ANISOU 5556 CZ TYR B 343 5053 4315 9121 -1211 -6 ATOM 5557 OH TYR B 343 -44.177 29.231 26.287 1.00 ANISOU 5557 OH TYR B 343 5289 4937 9483 -1610 -5 ATOM 5558 N PRO B 344 -50.486 26.216 28.466 1.00 ANISOU 5558 N PRO B 344 4927 3863 8281 127 -6 ATOM 5559 CA PRO B 344 -50.818 25.490 27.238 1.00 ANISOU 5559 CA PRO B 344 5321 4405 8540 87 -5 ATOM 5560 C PRO B 344 -50.414 26.285 26.009 1.00 ANISOU 5560 C PRO B 344 6648 5611 9826 -198 -5 ATOM 5561 O PRO B 344 -50.512 27.514 25.979 1.00 </pre>			
<input type="button" value="<"/> <input type="button" value=">"/>		<input type="button" value="^"/> <input type="button" value="▼"/>	

Figura 17: Aplicación que nos permite obtener las líneas de los $C\alpha$ a partir de la función WritePDB

Para certificar que el programa funcionaba correctamente y creaba un archivo válido, utilizamos YASARA y observamos que efectivamente el programa reconoce y representa los átomos que nos interesaban. Para ello es indispensable guardar el archivo con una extensión de tipo “.pdb”, el archivo en cuestión es “mod.pdb”.

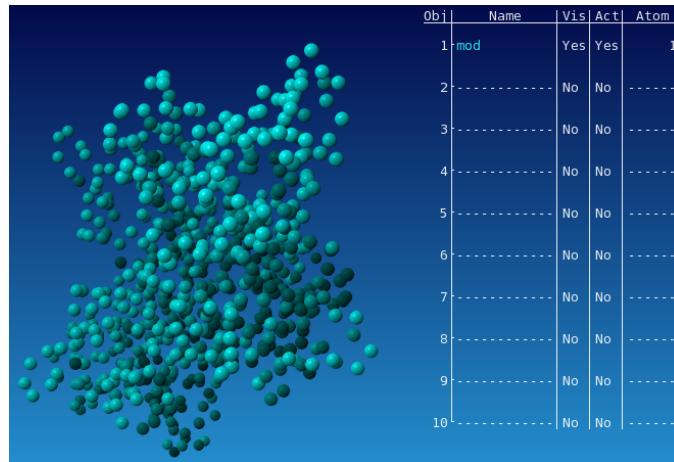


Figura 18: Figura obtenida a partir de la representación de los datos obtenidos a partir de WritePDB

4.3. Predicción estructural de proteínas

4.3.1. Ramachandran

Para hacer un diagrama de Ramachandran lo primero que debemos entender es qué son los ángulos de torsión de la proteína.

Los ángulos de torsión de la proteína son los ángulos llamados ϕ (de N a C α) y ψ (de C α a C), que oscilan entre 180 y -180 grados. Estos ángulos de la cadena principal (las laterales en este caso no tiene relevancia) son los que van a determinar la geometría y las propiedades físicoquímicas de la proteína. Adicionalmente, existe otro ángulo de torsión, denominado ω , pero este suele presentar valores establecidos por lo que podremos obviarlo para nuestros cálculos.

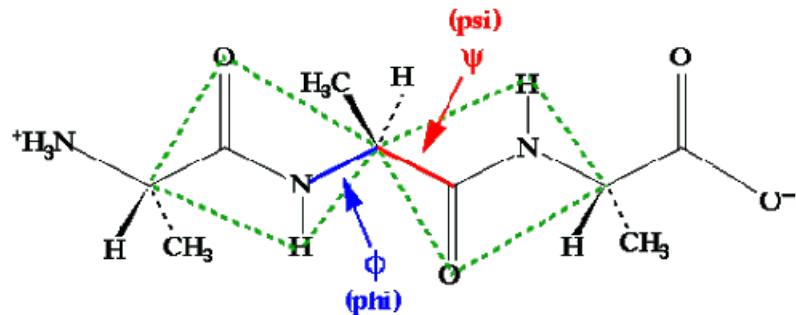
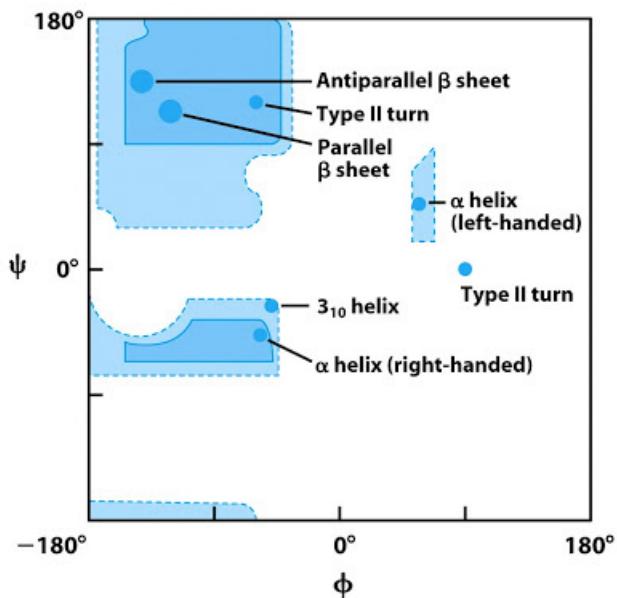


Figura 19: Representación de los ángulos de torsión

Los diagramas de Ramachandran son representaciones de los valores de los ángulos de torsión ϕ y ψ de los residuos de una proteína (a excepción del primero y el último, que no poseen ángulos ϕ y ψ , respectivamente). Estos diagramas tienen una gran importancia a la hora de determinar qué tipo de estructura secundaria es la dominante en la proteína, pues ya se conocen cuáles son las combinaciones de ángulos ϕ y ψ que dan lugar a hélices α , láminas β , giros y demás tipos de estructuras. Nuestra tarea consiste en crear un diagrama de Ramachandran de nuestra proteína de interés.



Para ello, crearemos los parámetros “phi” y “psi” en nuestra matriz de datos TResiduosPDB. El valor de estos ángulos será calculado a partir de los átomos de la cadena principal de ese residuo y del anterior o posterior, según trabajemos con ϕ o ψ .

Cabe destacar que por convenio se considera al ángulo de torsión como el que forman 4 átomos cuando los dos átomos centrales se encuentran en la intersección de los dos planos que contienen los 4 átomos, formando un diedro. Si observamos la estructura de forma que se superpongan los dos átomos centrales podremos ver el giro existente entre el primer átomo y el último, pudiendo encontrarse superpuestos (cis), en sentido opuesto (trans) y en todas las posiciones intermedias.

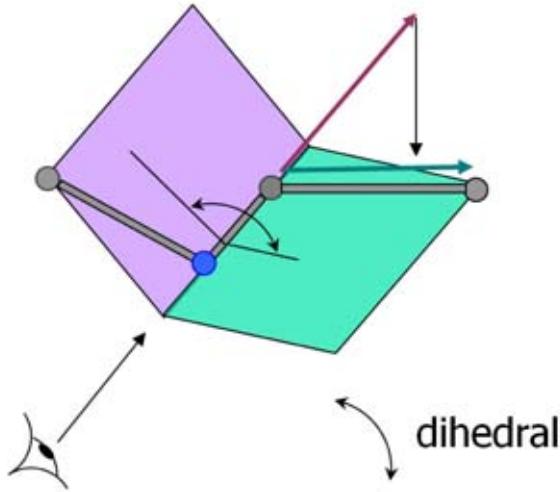


Figura 20: Perspectiva utilizada para entender los ángulos de torsión.

Para calcular los ángulos ϕ y ψ , necesitábamos las distancias entre los átomos que forman parte del ángulo diedro, para ello utilizamos el vector diferencia $\vec{v} = \vec{A} - \vec{B}$. Asimismo, para calcular el ángulo correspondiente nos valemos de la ecuación del cálculo del producto escalar entre dos vectores. Despejando obtenemos el ángulo de interés

$$\alpha = \arccos \frac{\vec{v}_1 \cdot \vec{v}_2}{\|\vec{v}_1\| \|\vec{v}_2\|} \quad (1)$$

De esta manera, aplicada a nuestro caso particular:

$$\alpha = \arccos \frac{(\vec{A} - \vec{B})(\vec{B} - \vec{C})}{\|\vec{A} - \vec{B}\| \|\vec{B} - \vec{C}\|} \quad (2)$$

Observamos que nuestra aplicación Ramachandran funciona correctamente con la proteína que hemos utilizado de prueba durante toda la asignatura “1AHI”, no obstante, con nuestra proteína asignada nos surgen errores, lo que podría deberse a que el archivo tipo pdb de la misma tiene diferencias respecto a los demás que hemos visto, en especial porque incluye un factor anisotrópico de temperatura y no representa todas las subunidades mencionadas.

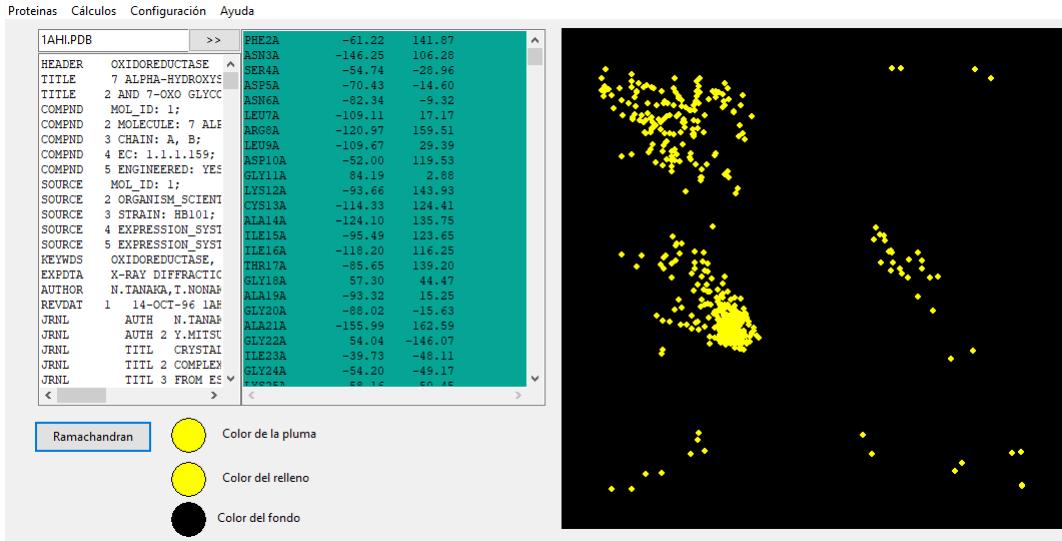


Figura 21: Perspectiva utilizada para entender los ángulos de torsión.

Para poder obtener los valores de los ángulos phi y psi recurrimos a una de las aplicaciones de “MolProbity”, desarrollada por el departamento de Bioquímica de la Universidad de Duke.

Residuo	ϕ	ψ
1	-72.92	136.32
2	-165.55	153.42
3	-64.36	-44.72
4	-66.80	-29.83
5	-72.99	-37.85
6	-61.97	-42.60
7	-67.09	-32.87
8	-59.94	-45.86
9	-64.99	-37.12
10	-66.35	-43.77

4.3.2. Coordenadas transformadas

Durante el progreso de la asignatura hemos creado diversas funciones que nos permiten realizar transformaciones en las coordenadas de los átomos de

nuestros TPuntos. Entre estas funciones se encuentran giros sobre los ejes X,Y,Z y translaciones.

En conjunto, entre ellas nos permiten realizar modificaciones en la perspectiva con la que vemos la proteína, y como ejemplo visualizaremos la proteína con un giro de 5 grados, lo que nos aporta una pequeña idea de cómo sería la imagen tridimensional de la proteína (considerando como 5 grados el ángulo entre los ojos del observador). En este caso, nos centraremos en la primera fenilalanina y los dos residuos siguientes, representando todos sus átomos correspondientes.

En otros programas ya hemos realizado representaciones y transformaciones, pero en este caso debemos adaptarlos al residuo requerido, tomando su átomo inicial y el átomo final del residuo n+2 en la cadena. Tomaremos sus coordenadas y crearemos una TTablaDatos para la representación.

El esquema lógico seguido es el siguiente: se carga un archivo y crea un TPDB, se seleccionan los datos del mismo y se representan los datos con y sin la modificación modificando TTablaDatos. Cabe mencionar que el valor seleccionado para el giro es de “0.087” pues es el valor correspondiente a 5 grados en radianes.

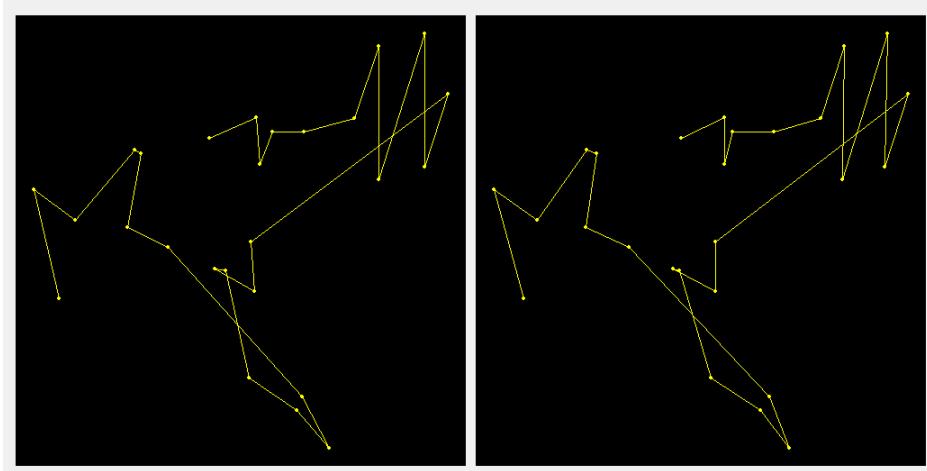


Figura 22: Imagen obtenida a partir del programa “Estereodiagrama” de nuestra proteína de interés. Las diferencias son ligeras debido a que el ángulo de giro es bastante reducido. Imagen modificada a la derecha

A pesar de ser un leve giro, se observa la profundidad de la posición de los átomos claramente en algunos átomos. Para asegurarnos de que habíamos realizado bien la aplicación y que el resultado sea correcto comparamos con imágenes obtenidas del software de visualización YASARA.

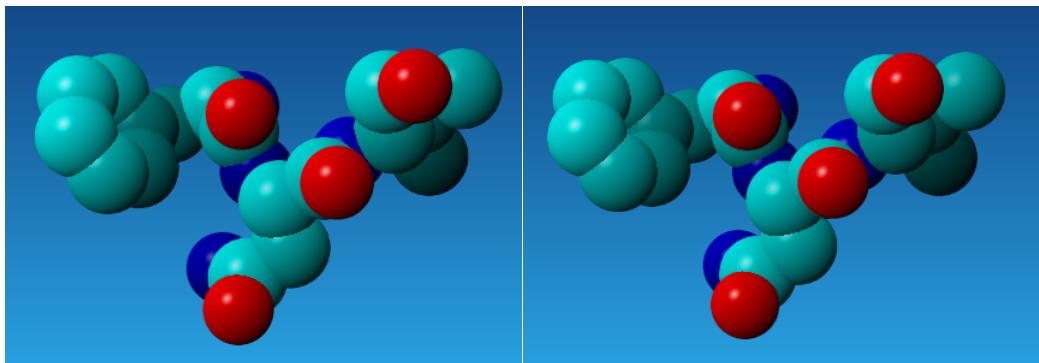


Figura 23: YASARA, imagen modificada a la derecha. El giro es más leve de ver si cabe debido al tamaño de las esferas correspondientes a cada átomo

4.3.3. Alinear Z

En esta tarea debemos rotar la proteína para visualizar desde el eje Z. No obstante, para ello hay que colocarla sobre ese eje y de nuevo, deshacer los cambios para obtener las coordenadas adecuadas.

Para ello lo primero que tenemos que hacer es crear una estructura que nos permite guardar la información relativa a las coordenadas de los átomos de una proteína, esta será llamada TPunto. Con ella, crearemos un array de puntos que denominaremos TPuntos. Ahora ya tendremos toda la información relativa a los átomos de la proteína, y podremos empezar a realizar los cálculos necesarios. En este caso tal y como nos pide el ejercicio solo tendremos en cuenta los C α .

A nivel de cálculos, deberemos realizar las siguientes transformaciones:

1. Translación de P1 al (0,0,0). Se traslada todo el objeto, incluido el P2. Recordemos que tenemos un eje que une ambos puntos.
2. Giro sobre OX hasta que P2 se encuentre en el plano XZ.
3. Giro sobre OY hasta que P2 esté en el eje Z
4. Giro del número de grados necesario sobre el eje OZ.
5. Inversa del paso 3
6. Inversa del paso 2
7. Inversa del paso 1

Es importante resaltar que en este tipo de transformaciones encadenadas de los giros el orden sí que es importante y es clave mantener una secuencia racional de cambios en las coordenadas de la proteína.

Para ello tenemos una función “alinearZ” que toma un TPuntos y lo devuelve modificado, a través de las siguientes transformaciones:

En primer lugar realizamos la translación mencionada.

Posteriormente, realizamos los cambios necesarios al P2 con las siguientes ecuaciones:

$$d1 = \sqrt{b^2 + c^2} \quad (3)$$

$$d2 = \sqrt{a^2 + b^2 + c^2} \quad (4)$$

$$\sin(\phi) = \frac{\text{cateto opuesto}}{d1} \quad (5)$$

$$\sin(\alpha) = \frac{\text{cateto contiguo}}{d2} \quad (6)$$

$$\phi = \arcsin(\sin(\phi)) \quad (7)$$

$$\alpha = \arcsin(\sin(\alpha)) \quad (8)$$

Finalmente, realizamos los giros correspondientes a los pasos 2 y 3 con los ángulos ϕ y α respectivamente. El resto de pasos se omiten para la problemática que se nos presenta. Es decir, obtenemos la representación desde la vista del eje OZ pero sin aplicar los pasos inversos que hemos mencionado anteriormente.

Respecto a nuestra proteína de interés, la Nsp15, los resultados de los 10 primeros $C\alpha$ fueron los que se muestran en la siguiente imagen:

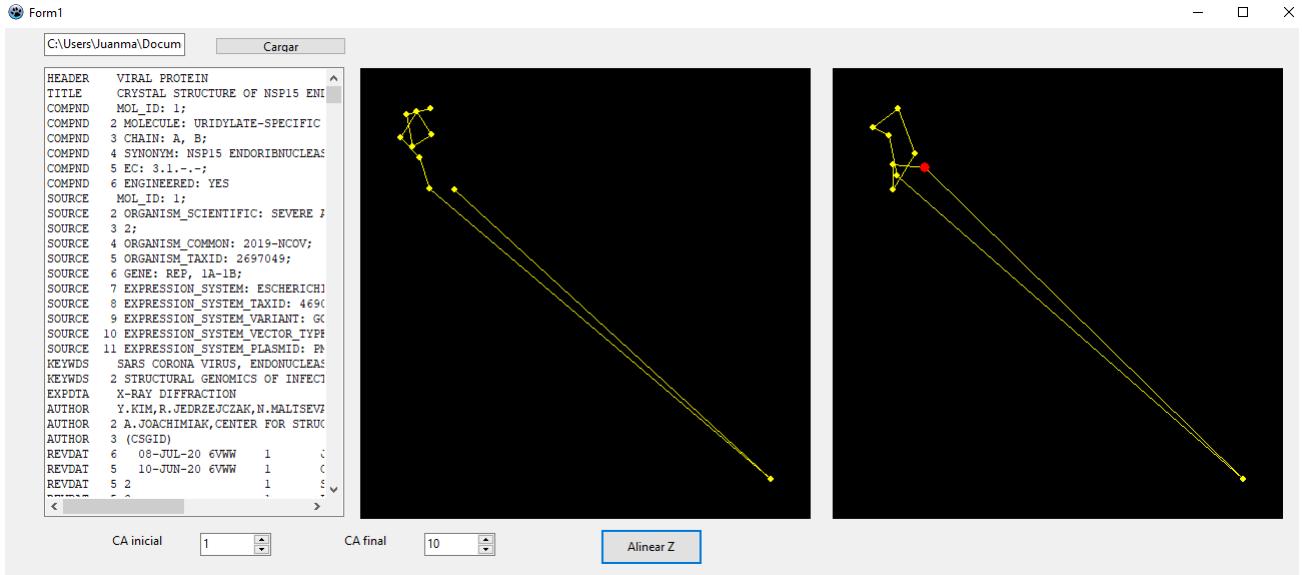


Figura 24: Imagen obtenida a partir del programa “alinearZ” de nuestra proteína de interés. Los átomos superpuestos se muestran en color rojo

Simplemente para comprobar que el programa seleccione correctamente los átomos, visualizamos con YASARA los primeros diez C_α , comprobando que la disposición es similar. Las diferencias se deben a que la vista no es exactamente la misma.

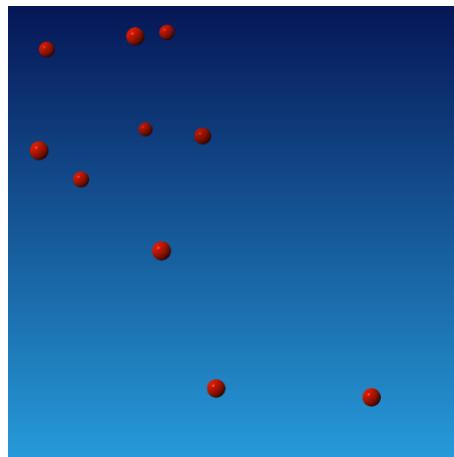


Figura 25: Imagen obtenida de YASARA

4.3.4. RMSD

En este ejercicio teníamos que estudiar la diferencia en distancia entre los átomos de las 3 primeras cisteínas de la proteína. Para ello, hemos seleccionado las 3 cisteínas con un bucle que buscaba el código de 3 letras “CYS” que presentan estos residuos.

Tras esto, hemos ido guardando los datos de las distancias entre los átomos en una matriz dinámica, donde los valores de la diagonal deben ser 0 pues un átomo consigo mismo no puede tener un valor diferente para este parámetro.

Tras esto, hemos calculado los “dij” como la suma de las distancias entre los átomos de cada cisteína (átomos obtenidos a partir de los TResiduoPDB), y por último hemos calculado las diferencias entre cada una de las cisteínas, es decir, nuestro RMSD.

$$RMSD = \sqrt{\frac{\sum_{i,j}^n (Dist1_{i-j} - Dist2_{i-j})^2}{n}}$$

Figura 26: Fórmula utilizada para el cálculo de los valores del RMSD de las cisteínas. Donde n=6 y las distancia son las obtenidas a partir del programa “RMSD”

Por ejemplo, obtenemos los valores correspondientes al RMSD de la siguiente proteína:

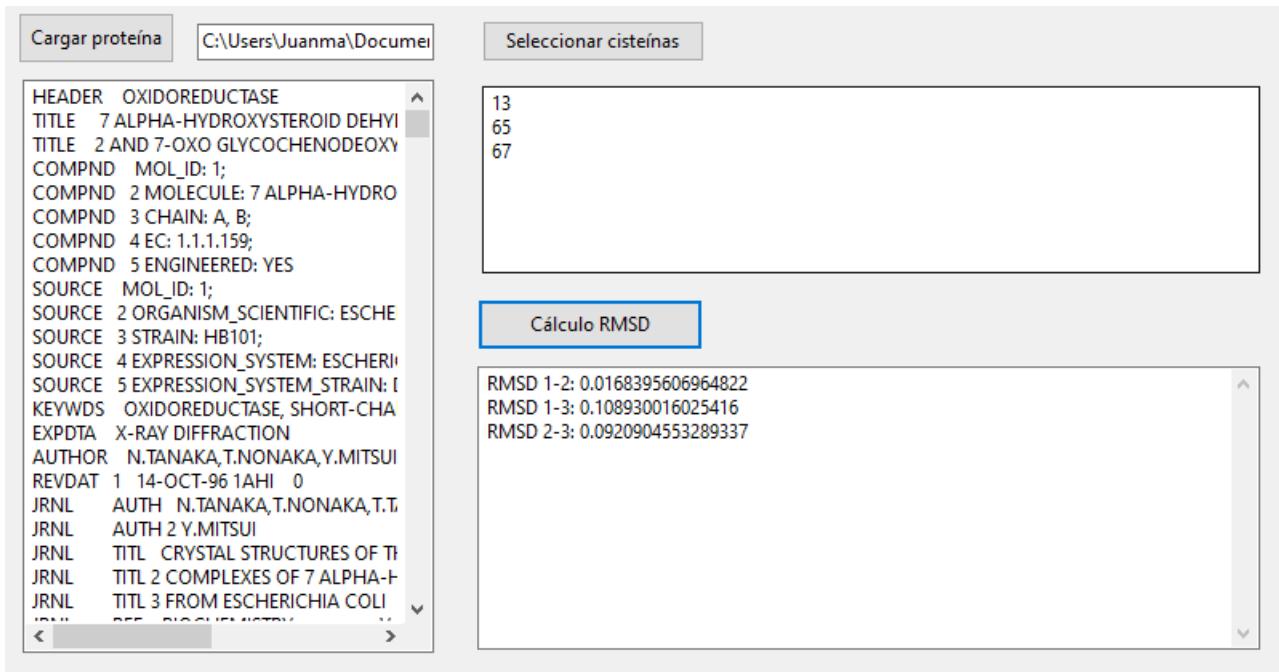


Figura 27: Fórmula utilizada para el cálculo de los valores del RMSD de las cisteínas. Donde n=6 y las distancia son las obtenidas a partir del programa “RMSD”

4.3.5. Mutación

En esta actividad se nos plantea la siguiente problemática: debemos estudiar cuál sería el efecto de una mutación en nuestra proteína. Más específicamente, una mutación missense de fenilalanina a tirosina ($\text{PHE} \rightarrow \text{TYR}$). He decidido escoger la mutación en la posición 303 de la cadena B de la proteína.

A efectos prácticos, la mutación de una fenilalanina a una tirosina únicamente conlleva la adición de un grupo OH en el extremo del anillo aromático. Para calcular la posición de este grupo, deberemos valernos de los vectores que determinan la posición del $C\zeta$ y del $C\beta$, que son, \overrightarrow{CZ} y \overrightarrow{CB} respectivamente, tomando como punto de partida el eje de coordenadas. A partir de ellos obtenemos el vector \overrightarrow{D} y por consiguiente, \overrightarrow{v} , así como la dirección y sentido de este último.

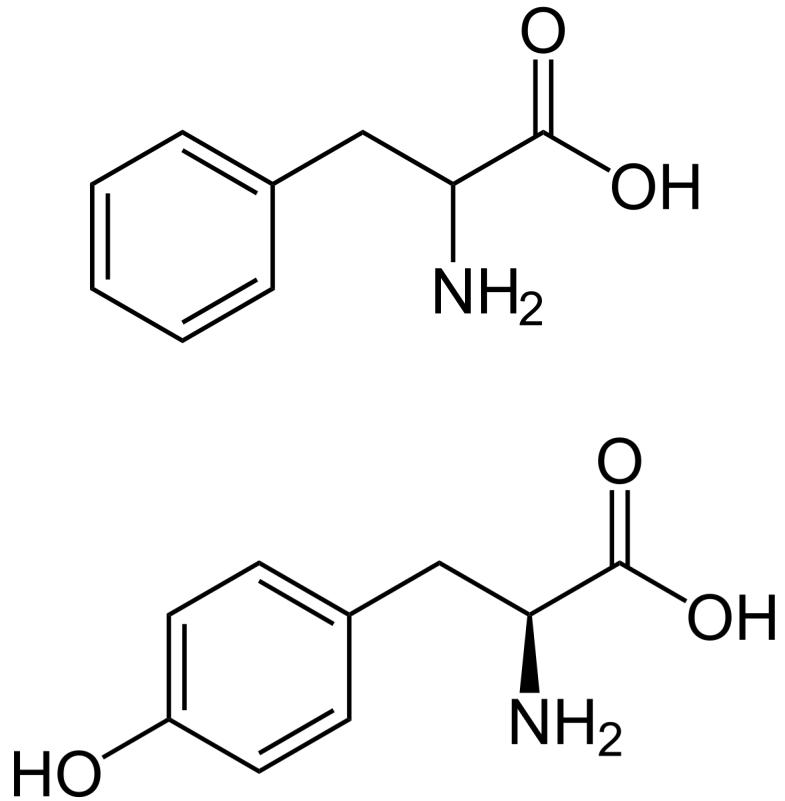


Figura 28: Comparación estructural provocada por la mutación de PHE a TYR

Hemos seleccionado tres tirosinas al azar de la proteína para estudiar como se comportaban sus vectores para obtener el módulo promedio $\|m\|$. Con ellos hemos determinado el módulo del vector. Esto nos ha permitido posteriormente con nuestra fenilalanina de interés calcular el vector unitario \vec{U} , y finalmente obtener el vector \vec{v} de interés, lo que nos permite conocer las coordenadas teóricas del grupo OH.

$$\|m\| = \vec{O} - \vec{C}\vec{Z} \quad (9)$$

$$\vec{D} = \vec{C}\vec{Z} - \vec{C}\vec{B} \quad (10)$$

$$\vec{U} = \frac{\vec{D}}{\|\vec{D}\|} \quad (11)$$

Con lo que finalmente obtenemos:

$$\vec{V} = \vec{U} * \|m\| \quad (12)$$

Finalmente, tomamos el \overrightarrow{CZ} de nuestro residuo, sumándole el vector \vec{V} a mutar y añadimos una línea ATOM nueva con las coordenadas obtenidas, siguiendo el mismo sentido que el esqueleto del anillo aromático.

	Referencia	coor X	coor Y	coor Z	vector
1	Cz	-65.609	47.528	-8.26	
	O	-66.946	47.659	-8.558	1.37605741
2	Cz	-45.273	29.062	27.1	
	O	-44.177	29.231	26.287	1.375044
3	Cz	-64.132	10.958	25.65	
	O	-65.065	10.303	26.42	1.37565039
					media vector 1.37558393
Fenilalanina		Modificación	coor X	coor Y	coor Z
PHE B 303		TYR B 303	-	-	-
		Cz	-51.405	24.943	36.003
		Cb	-54.889	22.642	36.955
		D	3.484	2.301	-0.952
		D/modulo_D	0.8135578	0.53731243	-0.222303969
		vector v	1.11911704	0.73911835	-0.305797768
		O	-50.285883	25.6821183	35.69720223

Figura 29: Matriz de cálculos realizados en el software Excel

ATOM	5222	CZ	PHE B 303	-51.405	24.943	36.003	1.00
44.32		C					
ANISOU	5222	CZ	PHE B 303	5506	4685	6650	
27	-1198	895	C				
ATOM	5222	OH	PHE B 303	-50.286	25.682	35.697	1.00
ATOM	5223	N	VAL B 304	-53.936	20.908	39.493	1.00
44.22		N					
ANISOU	5223	N	VAL B 304	6223	4414		

Figura 30: Línea ATOM creada

La mutación se ha producido en el dominio NendoU, responsable de la actividad catalítica, muy cerca además del centro activo, por lo que la carga introducida podría modificar la estructura del mismo y por ende, modificar su acción. Adicionalmente, la mutación afecta a un residuo que se encuentra en una zona hidrofóbica en el interior de la proteína, por lo que podría crear una mayor repulsión de cargas.

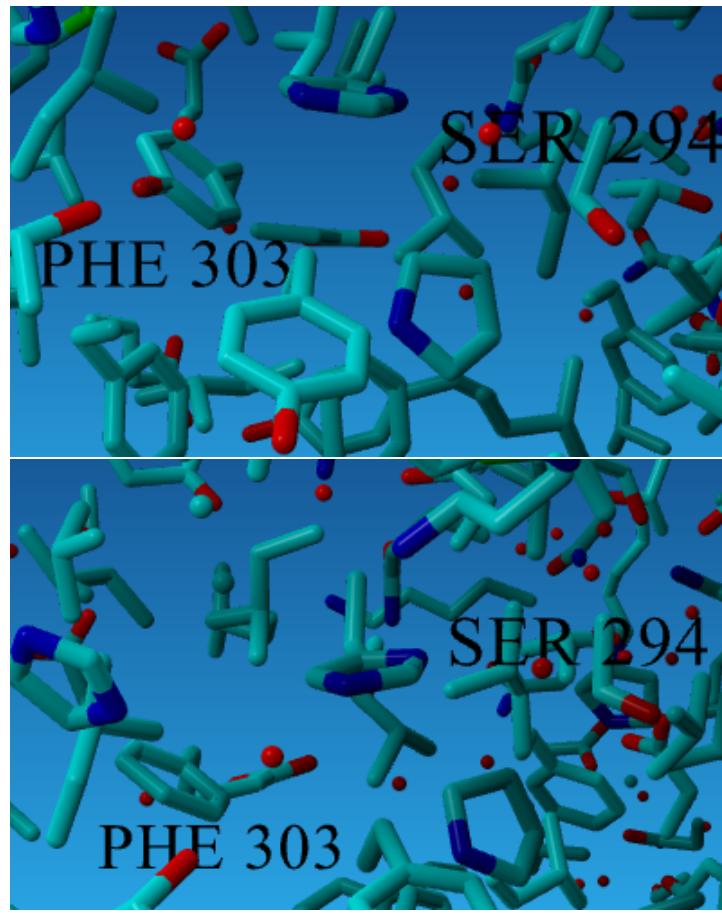


Figura 31: Visualización de la proteína mutada (arriba) y la proteína silvestre(abajo)

4.3.6. Enlaces disulfuro

En esta actividad hemos realizado un programa que recorre los átomos de la proteína buscando los átomos sulfuro cisteínas con ID “SG”, es decir, los que pueden dar lugar a la formación de puentes disulfuro.

En este ejercicio la distancia entre los átomos de sulfuro es esencial, pues de esta depende que se puedan crear este tipo de puentes, ya que solo a determinadas distancias se dan las energías correctas para su estabilización. Además, mientras unas proteínas pueden tener una gran cantidad de este tipo de enlaces, otras pueden carecer completamente de ellos.

Por ello, realizamos previamente al estudio de nuestra proteína un screening con otras que utilizamos de base para encontrar la distancia en Angstrom adecuada para que se den los puentes disulfuro.

Por ejemplo, 1HAI y 1BHS no contiene ninguno, mientras que otras como 2ZUP contiene apenas 2 y 2QWQ únicamente 1, y otras como 1AFD tiene hasta 6 enlaces. Del muestreo realizado deducimos que la distancia correcta para el enlace disulfuro se encuentra aproximadamente en 2.05 Angstrom, si bien podemos dar por válidas distancias un poco mayores para asegurarnos de no perder ningún puente disulfuro.

\Documents\Ingeniería de proteínas\2ZUP.pdb >>>

HELIX	7	7	SER A	128	VAL A	145	1					
HELIX	8	8	ASN A	170	GLU A	187	1					
HELIX	9	9	ALA B	14	LEU B	23	1					
HELIX	10	10	LEU B	25	VAL B	35	1					
HELIX	11	11	VAL B	42	GLY B	53	1					
HELIX	12	12	TIR B	71	TIR B	79	1					
HELIX	13	13	SER B	80	LEU B	94	1					
HELIX	14	14	PRO B	115	VAL B	120	1					
HELIX	15	15	PRO B	143	ILE B	151	1					
HELIX	16	16	ILE B	151	ILE B	162	1					
SHEET	1	A	3	TIR A	9	THR A	11	0				
SHEET	2	A	3	TIR A	159	LEU A	161	-1	O	GLN A	160	N
SHEET	3	A	3	PHE A	154	VAL A	155	-1	O	VAL A	155	N
SHEET	1	B	2	VAL A	22	GLU A	24	0				
SHEET	2	B	2	MET A	56	LYS A	58	1	O	THR A	57	N
SSBOND	1	CYS A	30	CYS B	104							
SSBOND	2	CYS B	41	CYS B	44							
LINK			NE2 HIS A	41					ZN	ZN A	190	
CISPEP	1	LEU A	82	GLY A	83				0		10.0	
CISPEP	2	VAL A	150	PRO A	151				0		4.1	
SITE	1	AC1	10	ALA B	29	GLN B	33	LYS B	39	PRO B	40	

Umbra: 3.80

Predicción de enlaces disulfuro

Cisteína 1	Cisteína 2	Distancia
CYS30A	CYS104B	2.04
CYS41B	CYS44B	2.03

C:\Users\Juanma\Documents\Ingeniería de pro >>>

SHEET	5	B	5	ASN A	168	ASN A	174	1	O	LEU A	170	N
SHEET	1	C	3	ARG A	49	ILE A	51	0				
SHEET	2	C	3	VAL A	42	PHE A	44	-1	N	ALA A	43	O
SHEET	3	C	3	THR A	66	VAL A	67	-1	O	VAL A	67	N
SHEET	1	D	3	MET A	93	ASP A	97	0				
SHEET	2	D	3	ARG A	100	TYR A	107	-1	O	GLN A	104	N
SHEET	3	D	3	GLU A	110	PHE A	114	-1	O	LYS A	112	N
SHEET	1	E	4	ILE A	216	ASP A	225	0				
SHEET	2	E	4	PHE A	205	GLU A	213	-1	N	VAL A	207	O
SHEET	3	E	4	ARG A	193	LEU A	200	-1	N	ILE A	197	O
SHEET	4	E	4	ASP A	333	VAL A	337	1	O	ASP A	333	N
SHEET	1	F	2	GLN A	279	TYR A	288	0				
SHEET	2	F	2	ILE A	291	THR A	298	-1	O	PHE A	293	N
SSBOND	1	CYS A	171	CYS B	876							
SITE	1	AC1	7	GLY A	12	THR A	13	LYS A	71	GLU A	1	
SITE	2	AC1	7	THR A	204	ADF A	487	HOH A	3180			
SITE	1	AC2	20	THR A	13	THR A	14	TYR A	15	GLY A	2	
SITE	2	AC2	20	GLY A	202	GLY A	230	GLU A	268	LYS A	2	
SITE	3	AC2	20	ARG A	272	SER A	275	GLY A	338	GLY A	3	
SITE	4	AC2	20	SER A	340	ARG A	342	ILE A	343	ASP A	3	
SITE	5	AC2	20	PO4	488	GOL A	3148	HOB A	3180	HOB A	32	

Umbra: 3.80

Predicción de enlaces disulfuro

Cisteína 1	Cisteína 2	Distancia
CYS171A	CYS876B	2.05

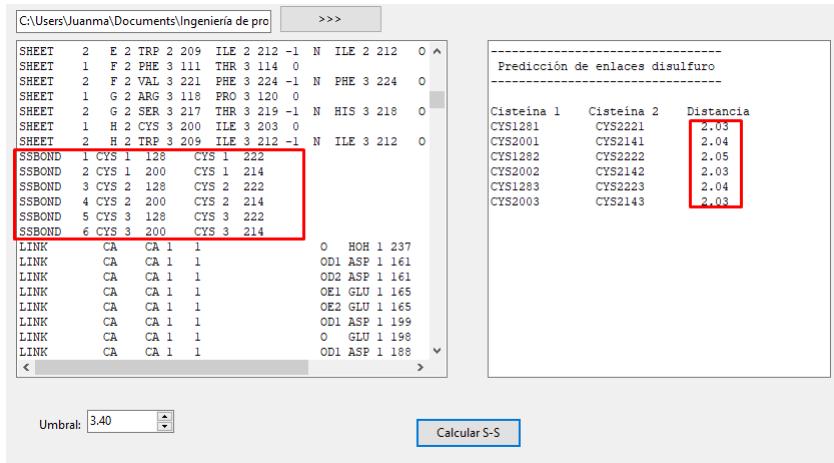


Figura 32: Imágenes tomadas de la aplicación “Enlaces SS” para tomar referencias de cara al estudio de los enlaces disulfuro en nuestra proteína de interés Nsp15

Para asegurarnos que nuestra predicción era correcta, comprobamos que no existía ningún apartado para enlaces disulfuro en el archivo PDB de la proteína, y además en YASARA tampoco encontramos ninguno.

4.3.7. Anfipatía-hidrofobicidad

La hidrofobicidad es una fuerza dominante en la formación de la estructura tridimensional plegada de la proteína. Básicamente se refiere a la fuerza que controla que las regiones apolares no estén expuestas al solvente polar, impidiendo la formación de clatratos, estructuras muy ordenadas y desfavorables a nivel energético. Por ende, se formarán interacciones entre moléculas hidrofóbicas, lo que redundaría en una disminución neta de la energía del sistema. Por tanto, hablamos de una fuerza esencialmente entrópica.

$$\Delta G = \Delta H - T\Delta S \quad (13)$$

Teniendo en cuenta que todos los residuos comparten su esqueleto covalente, la hidrofobicidad vendrá determinada por el carácter hidrofóbico de las cadenas laterales. No obstante, es muy complicado obtener un valor directo de la hidrofobicidad de cada residuo, pues se ve influenciado por el ambiente

en el que se encuentra.

Por ello, existen diversas escalas hidrofóbicas, donde podemos destacar alguna como es el caso de la de “doolittle”. Habrá que cargar esta escala en nuestra aplicación para poder llevar a cabo el análisis correspondiente.

Desgraciadamente, obtener el valor de hidrofobicidad correspondiente a cada residuo es una tarea muy compleja, por lo que la estrategia a seguir será la de asignar a un residuo el valor promedio de la hidrofobicidad obtenida de una ventana de residuos contiguos a través del “moving average”. De esta manera podremos reducir el ruido y mejorar la fiabilidad de nuestros datos. Esto ha sido denominado “método de suavizado de Savinski-Golay”.

Otro factor importante que mediremos en este programa es la denominada anfipatía, que se refiere a la hidrofobicidad medida en la distribución de la estructura secundaria de la proteína. Utilizamos dos métodos para el estudio de dicha anfipatía:

1. Eisenberg. Este método se basa en el estudio del momento hidrofóbico midiendo magnitudes físicas, estableciéndolo la hidrofobicidad total como una suma de vectores

$$\mu_H = \sqrt{(\sum H_n \sin(\delta n))^2 + (\sum H_n \cos(\delta n))^2} \quad (14)$$

2. Stroud. Este método se basa en el uso de magnitudes matemáticas (“Fourier Power Spectrum”). De la transformada de Fourier se obtiene la amplitud respecto a la frecuencia del momento armónico representado.

$$I(k, v) = \left(\sum_{j=k-n}^{k+n} (h_j - \bar{h}(k)) \sin(jv) \right)^2 + \left(\sum_{j=k-n}^{k+n} (h_j - \bar{h}(k)) \cos(jv) \right)^2 \quad (15)$$

En lo referido a nuestro programa, la estrategia seguida será la siguiente.

Necesitaremos 3 inputs: la proteína elegida, la semiventana y la escala. Además, podremos seleccionar los residuos en los que mediremos la hidrofobicidad y anfipatía. Cargaremos los datos obtenidos a partir de las ecuaciones

correspondientes en una estructura TTablaDatos, y con esta podremos representar los perfiles hidrofóbicos mediante un plot.

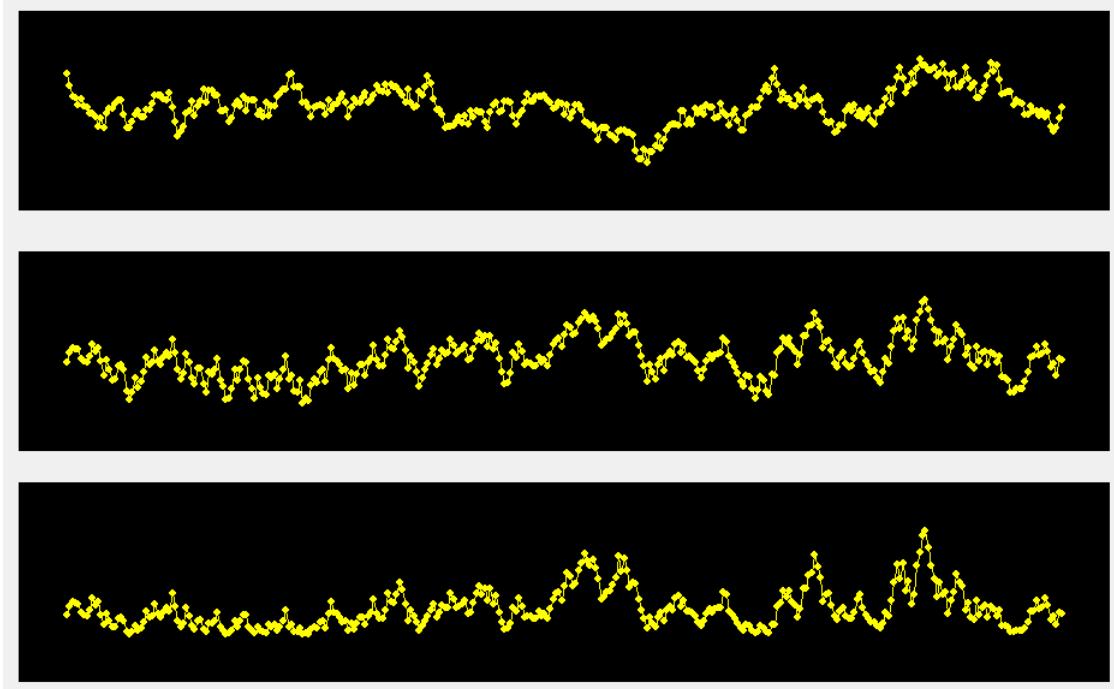


Figura 33: Imágenes correspondientes a los perfiles de hidrofobicidad, anfipatía por Eisenberg y anfipatía por Stroud

Como podemos observar los datos obtenidos en las gráficas son similares, lo que realza nuestra confianza en el programa. Podemos inferir que los puntos altos de las gráficas de anfipatía se corresponden con hélices α , pues la anfipatía se define como la lateralidad de la hidrofobicidad, lo que sucede en cada giro de la hélice. Por su parte los “llanos” de la gráfica se corresponden con láminas β , que tienen una menor anfipatía por su disposición. No obstante, esto puede llevarnos a falsas interpretaciones, y es muy importante optimizar en cada caso la semiventana utilizada.

Referencias

- Batool, A., Bibi, N., Amin, F., y Kamal, M. (2021). Drug designing against nsp15 of sars-cov2 via high throughput computational screening and structural dynamics approach. *European Journal of Pharmacology*, 892.
- Kim, Y., Jedrzejczak, R., Maltseva, N., Wilamowski, M., Endres, M., Godzik, A., ... Joachimiak, A. (2020). Crystal structure of nsp15 endoribonuclease nendou from sars-cov-2. *Protein Science*, 29(7), 1596-1605.
- Kim, Y., Wower, J., Maltseva, N., Chang, C., Jedrzejczak, R., Wilamowski, M., ... Joachimiak, A. (2020). Tipiracil binds to uridine site and inhibits nsp15 endoribonuclease nendou from sars-cov-2. *bioRxiv*.
- Kimura, H., Kurusu, H., Sada, M., Kurai, D., Murakami, K., Kamitani, W., ... Ryo, A. (2020, 08). Molecular pharmacology of ciclesonide against SARS-CoV-2. *J Allergy Clin Immunol*, 146(2), 330–331.
- Kumar, R., Verma, H., Singlhi, N., Sood, U., Gupta, V., Singh, M., ... Lal, R. (2020). Comparative genomic analysis of rapidly evolving sars-cov-2 reveals mosaic pattern of phylogeographical distribution. *mSystems*, 5(4).
- Kumar, S., Kashyap, P., Chowdhury, S., Kumar, S., Panwar, A., y Kumar, A. (2020). Identification of phytochemicals as potential therapeutic agents that binds to nsp15 protein target of coronavirus (sars-cov-2) that are capable of inhibiting virus replication. *Phytomedicine*, 153317.
- Li, Y., Yang, X., Wang, N., Wang, H., Yin, B., Yang, X., y Jiang, W. (2020). Gc usage of sars-cov-2 genes might adapt to the environment of human lung expressed genes. *Molecular Genetics and Genomics*, 295(6), 1537-1546.
- Matsuyama, S., Kawase, M., Nao, N., Shirato, K., Ujike, M., Kamitani, W., ... Fukushi, S. (2020). The inhaled corticosteroid ciclesonide blocks coronavirus rna replication by targeting viral nsp15. *bioRxiv*.
- Ortiz-Alcantara, Bhardwaj, Palaninathan, Frieman, Baric, y Kao. (2010). Small molecule inhibitors of the sars-cov nsp15 endoribonuclease. *Virus Adaptation and Treatment*, 2:125-133.
- Pillon, M., Frazier, M., Dillard, L., Williams, J., Kocaman, S., Krahn, J., ... Stanley, R. (2021). Cryo-em structures of the sars-cov-2 endoribonuclease nsp15 reveal insight into nuclease specificity and dynamics. *Nature Communications*, 12(1).
- Savale, R., Bhowmick, S., Osman, S., Alasmary, F., Almutairi, T., Abdullah,

- D., ... Islam, M. (2021). Pharmacoinformatics approach based identification of potential nsp15 endoribonuclease modulators for sars-cov-2 inhibition. *Archives of Biochemistry and Biophysics*, 700.
- Sharma, A., Goyal, S., Yadav, A. K., Kumar, P., y Gupta, L. (2020). In-silico screening of plant-derived antivirals against main protease, 3clpro and endoribonuclease, nsp15 proteins of sars-cov-2. *Journal of Biomolecular Structure and Dynamics*, 0(0), 1-15.
- Sinha, S. K., Shakya, A., Prasad, S. K., Singh, S., Gurav, N. S., Prasad, R. S., y Gurav, S. S. (2020). An in-silico evaluation of different saikosaponins for their potency against sars-cov-2 using nsp15 and fusion spike glycoprotein as targets. *Journal of Biomolecular Structure and Dynamics*, 0(0), 1-12.
- Umar, H., Siraj, B., Ajayi, A., Jimoh, T., y Chukwuemeka, P. (2021). Molecular docking studies of some selected gallic acid derivatives against five non-structural proteins of novel coronavirus. *Journal of Genetic Engineering and Biotechnology*, 19(1).