



Representació tridimensional de les proteïnes amb Panda 3D

Josep Maria Escofet Figueras

Máster universitari en Bioinformática y Bioestadística. Treball final de Màster.

Brian Jiménez García.

Juny -2018

A) Creative Commons:



Esta obra está sujeta a una licencia de
Reconocimiento-NoComercial-
SinObraDerivada [3.0 España de Creative
Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

B) GNU Free Documentation License (GNU FDL)

Copyright © 2018 Josep Maria Escofet Figueras

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

C) Copyright

© Josep Maria Escofet Figueras

Reservados todos los derechos. Está prohibido la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la impresión, la reprografía, el microfilme, el tratamiento informático o cualquier otro sistema, así como la distribución de ejemplares mediante alquiler y préstamo, sin la autorización escrita del autor o de los límites que autorice la Ley de Propiedad Intelectual.

Titulo del Trabajo:	Representació tridimensional de les proteïnes:
Nombre del autor:	Josep Maria Escofet Figueras
Nombre del consultor/a:	Brian Jiménez García
Nombre del PRA:	Brian Jiménez García
Fecha de entrega (mm/aaaa):	6/2018
Titulació:	Màster Universitari de Bioestadística i Bioinformàtica
Àrea del Trabajo Final:	M0.177 - TFM-Estadística y Bioinformàtica
Idioma del trabajo:	Català
Palabras clave	Proteïnes, Representació tridimensional, Estructura de les proteïnes
Resumen del Trabajo (màximo 250 palabras): Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.	
<p>El projecte final del màster tracta el tema de la representació de l'estructura tridimensional de les proteïnes. La importància de l'estudi es basa en el fet que la funció de les proteïnes, òbviament, depèn de la seva composició, pel que fa als aminoàcids que la integren i la seqüència en què estan organitzats dins de la macromolècula, però també la seva forma tridimensional. Amb diversos hèlixs, làmines o girs , ja que això implica la possibilitat de connectar amb altres molècules o fins i tot amb àcids nucleics</p> <p>L'objectiu és la representació en 3D de la proteïna funcional a partir de les dades contingudes en la base de dades del Protein Data Bank en forma de fitxer .pdb. tot i que actualment hi ha multitud de visors de proteïnes, la idea del treball era fer un visor per python que no fos de pagament i que no utilitzes C++ ni java Java ni tingués que utilitzar la programació amb OpenGL . En el seu lloc s'utilitzaran les llibreries Panda3D i Matplotlib. El llenguatge escollit és el python 2.7.15</p> <p>L'entorn de desenvolupament utilitzat ha sigut l'editor IDE PyCharm</p> <p>El projecte inclou varis tipus de representació i es poden utilitzar diferents fitxers .pdb que s'han inclòs en el projecte. Les representacions són les formes superficials a partir del dibuix de cadascun dels àtoms segon el radi del element. L'altre forma utilitzada és el ball & stick a partir de la representació dels carbonis α i la unió entre àtoms alfa mitjançant unions lineals i finalment un pseudo backbone</p>	

Abstract (in English, 250 words or less):

The final project of the master's degree deals with the subject of the representation of the three-dimensional structure of the proteins. The importance of the study is based on the fact that the function of the proteins obviously depends on its composition, with respect to the amino acids that integrate it and the sequence in which they are organized within the macromolecule, but also its three-dimensional shape. With several helices, sheets or turns, since this implies the possibility of connecting with other molecules or even with nucleic acids.

The objective is the 3D representation of the functional protein, part of the data contained in the database of the Protein Data Bank in the form of a .pdb file. Although there are currently a multitude of protein viewers, the idea of the work was to make a Python viewer that was not paid and did not use C++ or Java. It did not even have to use OpenGL programming. Instead, the Panda3D and Matplotlib libraries will be used. The chosen language is Python 2.7.15.

The development environment used has been the PyCharm IDE editor.

The project includes several types of representation and different .pdb files that have been included in the project can be used. The representations are the superficial forms from the drawing of each one of the atoms according to the radius of the element. The other form used is the ball & stick from the representation of the α carbons and the union between atoms alpha through linear joints and finally a pseudo backbone.

The operation is by line of orders.

CONTENIDO

<u>1. INTRODUCCIÓ.....</u>	<u>11</u>
1.1 CONTEXT Y JUSTIFICACIÓ DEL TREBALL	11
1.2 OBJECTIUS DEL TREBALL.....	12
1.3 ENFOCAMENT I MÈTODE SEGUIT	13
1.3.1 CLASSES ALFANUMÈRIQUES	15
1.3.2 CLASSES GRÀFIQUES	15
1.4 PLANIFICACIÓ DEL TREBALL.....	18
1.5 BREU SUMARI DE PRODUCTES OBTINGUTS.....	20
1.6 BREU SUMARI DELS PRODUCTES OBTINGUTS.....	22
1.7 BREU DESCRIPCIÓ DELS ALTRES CAPÍTOLS DE LA MEMÒRIA	22
<u>2. ESTRUCTURA DE LES PROTEINES.....</u>	<u>22</u>
2.1 ESTRUCTURA PRIMÀRIA DE LA PROTEÍNA	24
2.2 ESTRUCTURA SECUNDARIA DE LAS PROTEINAS.....	25
2.3 ESTRUCTURA TERCIÀRIA DE LES PROTEÏNES.....	27
2.4 ESTRUCTURA QUATERNÀRIA DE LES PROTEÏNES	28
2.5 FUNCIONS DE LES PROTEÏNES	29
<u>3. ESTAT DE L'ART.....</u>	<u>30</u>
3.1 INTRODUCCIÓ.....	30
3.2 CC4MG	31
3.3 PYMOL	32
3.4 JMOL I JSMOL	33
3.4.1 JENA3D	34
3.5 MATLAB.....	35
3.6 EPMV	36
3.7 PMV (PYTHON MOLECULAR VIEWER)	37
<u>4. EL PROTEIN DATA BANK I EL TRACTAMENT INFORMÀTIC DELS FITXERS PDB</u>	<u>38</u>

4.1	TIPUS DE REGISTRES	38
4.1.1	UNA INFORMACIO – UNA LINIA.....	38
4.1.2	UNA SOLA INFORMACIO, EN UNA O DIVERSES LÍNIES	38
4.1.3	VÀRIES VEGADES, UNA LÍNIA	40
4.1.4	DIVERSES VEGADES, DIVERSES LÍNIES	41
4.1.5	ALTRES:.....	41
4.2	FORMA DE LA TAULA	41
4.2.1	LA FITXA GENERAL QUE TREU L'APLICACIÓ	45
4.3	ESTRUCTURA DE LES DADES	47
5.	<u>ESTRUCTURA DEL PROGRAMA</u>	52
5.1	CODI.ESTRUCTURA	53
5.1.1	CLASSE ENTITAT.....	53
5.1.2	FUNCIÓ MAXMINS	55
5.2	CODI.ATRIBUTS	55
5.2.1	CLASSE COLORS	55
5.2.2	ATRIBUTS	56
5.3	CODI.DADES	57
5.3.1	CLASSE DADES	57
5.4	PLOT2D.PROTEINA2D	59
5.4.1	CLASSE PLOTS.....	59
5.5	CODI.PLOT3D.PROTEINA3D	62
5.5.1	CLASSE PROTEINES	62
5.5.2	METODES	63
5.5.3	EXEMPLES GRÀFICS.....	67
5.5.4	SOBRE LA RENDERITZACIÓ.....	71
6.	<u>RESTA DE MODULS</u>	73
6.1	.GITIGNORE	73
6.2	AMINOACIDS.PY	73
6.3	COLORS.PY	73
6.4	FITXERS	74
6.5	DADES_BAT.PY	75

6.6	OPCIONES.PY	75
6.7	VISOR2D	75
6.8	VISOR3D	75
7.	<u>CONCLUSIONS</u>	76
8.	<u>GLOSARI</u>	78
9.	<u>BIBLIOGRAFIA</u>	79
10.	<u>ANNEXS</u>	80

Taula d'il·lustracions

	<i>Figura 1. Esquema del flux de treball en l'anàlisi de les proteïnes.....</i>	13
	<i>Figura 2 , Entorn IDE PyCharm.....</i>	14
	<i>Figura 3. Finestra del projecte amb els mòduls i fitxers auxiliars.....</i>	14
	<i>Figura 4. Exemple scatterplot de les tres molècules de la proteïna (SERT) 5I6X.pdb.....</i>	16
	<i>Figura 5. Mateixa proteïna representada anteriorment però visualitzada amb panda3D.....</i>	18
	<i>Figura 6. Paleta de colors</i>	21
	<i>Figura 7. Esquema de la formació de les proteïnes</i>	23
	<i>Figura 8. A la dreta el patró base d'un aminoàcid on es pot veure els dos extrem amb el grup carboxílic i a l'altre extrem el grup amino, essent el braç R que era substituint pels elements que diferencien els diferents aminoàcids. A l'esquerra està la formació de l'enllaç peptídic unint el radical amino d'un aminoàcid amb el carboxílic de l'altra alliberant el molècula d'aigua.....</i>	23
	<i>Figura 9.- Els 20 aminoàcids essencials http://deproteinas.com/los-20- aminoacidos-esenciales</i>	23
	<i>Figura 10.- Estructura primària de la proteïna http://umaquimicairresistivel.blogspot.com.es/2011/07/nossa-quimica.html. 24</i>	24

<i>Figura 11.- Plegament helix α a partir d'enllaços formats per ponts d'hidrogen.</i> <i>http://delnutrientealadieta.com/2016/02/03/modulo-5-1-naturaleza-tipos-y-propiedades-nutricionales-de-las-proteinas/19/.....</i>	25
<i>Figura 12 .- Plegament làmines β.</i> <i>http://delnutrientealadieta.com/2016/02/03/modulo-5-1-naturaleza-tipos-y-propiedades-nutricionales-de-las-proteinas/19/.....</i>	26
<i>Figura 13 .- Plegament terciari.</i> <i>http://delnutrientealadieta.com/2016/02/03/modulo-5-1-naturaleza-tipos-y-propiedades-nutricionales-de-las-proteinas/19/.....</i>	28
<i>Figura 14.- Estructura quaternària de la proteïna . Exemple de l'hemoglobina.</i> <i>Life: the science of biology , 7^a Edició, 2004.....</i>	28
<i>Figura 15- Representació amb CC$\\$mg supressor de tumors p53 complexat amb ADN.....</i>	31
<i>Figura 16- Representació amb PyMol supressor de tumors p53 complexat amb ADN.....</i>	32
<i>Figura 17- Representació amb Jmol supressor de tumors p53 complexat amb ADN.....</i>	33
<i>Figura 18- Representació amb Jena3d supressor de tumors p53 complexat amb ADN.....</i>	34
<i>Figura 19. Proteïna 1TUP realitzat amb Matlab, com es `pot veure el motor Jmol</i> <i>.....</i>	36
<i>Figura 20.- Estructura de les dades</i>	47
<i>Figura 21. Fòrmules dels aminoàcids Serina i Valina.....</i>	49
<i>Figura 22.- Fòrmula del Triptòfan.....</i>	50
<i>Figura 23.- Tros d'estructura secundària.</i> <i>https://maestradelia.files.wordpress.com/2012/11/ch3f6.jpg</i>	51
<i>Figura 24. Proteïna 1TUP per àtoms i molècules.....</i>	60
<i>Figura 25. Proteïna 1TUP per residus i molècules</i>	60
<i>• Figura 26. Proteïna 1TUP per línies(aminoàcids) i molècules</i>	61
<i>Figura 27. Proteïna 1TUP per línies i tipus aminoacids</i>	61
<i>Figura 28. Proteïna 1TUP diferents vistes.....</i>	62
<i>Figura 29.- .Representació del complex proteic segons color aminoàcids</i>	67

<i>Figura 30.- Representació dels elements externs a la proteïna.....</i>	<i>68</i>
<i>Figura 31.- Representació dels elements externs a la proteïna.....</i>	<i>68</i>
<i>Figura 32 Representació ubicació d'un aminoàcid concret</i>	<i>69</i>
<i>Figura 33 .- Representació del complex per molècules.....</i>	<i>69</i>
<i>Figura 34.- Representació ball & stick els acolorit per tipus aminoàcid</i>	<i>70</i>
<i>Figura 35.- Representació tipus backbone.....</i>	<i>70</i>
<i>Figura 36. Renderització segons plataforma i nombre de residus en el cas de representació atòmica</i>	<i>72</i>
<i>Figura 37.- Representació del colàgen per molècules en els dos sistemes.....</i>	<i>76</i>

1. INTRODUCCIÓ

1.1 CONTEXT Y JUSTIFICACIÓ DEL TREBALL

Descripció general: Explicar grans trets en què consistirà el TFM. Aquest punt no s'ha de confondre amb una introducció.

Aquest treball s'emmarca en una part de la biologia que cada vegada va agafant més importància, l'estructura tridimensional de les proteïnes. En part gràcies a les noves o a la millora de les antigues tècniques per estudiar les proteïnes així actualment tenim¹:

- Difracció de fibres, usada per a l'anàlisi de fibres amb patrons repetitius.
- Microscòpia electrònica, per a l'estudi de grans complexos moleculars. La microscòpia criar-electrònica (va criar-EM) és l'aproximació més prometedora en l'actualitat (Callaway, 2015).
- Cristal·lografia de raigs-X, aplicable a totes les macromolècules, permet obtenir les descripcions estructurals estàtiques de més qualitat a partir de vidres. Un bon llibre per llegir sobre això és Rhodes (2000). Les fonts de radiació de raigs X de làsers d'electrons lliures (XFEL) permeten estudiar cristalls petits i proteïnes de membrana (Marx, 2014).
- Dispersió de raigs X a baixos angles (SAXS), que no requereix l'obtenció de cristalls i s'ha utilitzat per a l'estudi de proteïnes desordenades.
- Anàlisi d'espectres de ressonància magnètica nuclear (NMR), aplicable a totes les macromolècules, permet estudiar comportaments dinàmics, com unions entre molècules i moviments moleculars (Jiang & Kalodimos, 2017). La construcció de models de proteïnes a partir de dades de NMR s'ha aconseguit automatitzar en casos senzills (Liu et al., 2017; Rosato et al., 2012). Un llibre de referència sobre aquesta metodologia és Cavanagh et al. (2005).

¹ Veure Bruno Contreras-Moreira <http://www.eead.csic.es/compbio>

- Anàlisi d'espectres de dicroisme circular (CD), que permet estimar de manera senzilla al laboratori el percentatge de residus d'una seqüència que adquireixen estructura secundària. Els espectres obtinguts es poden comparar amb els d'estructures conegudes (Mavridis & Janes, 2017).
- Espectrometria de masses de molècules entrecruades, que ha permès obtenir informació estructural per modelar complexos de proteïnes (Rappsilber, 2011).
- La ultra seqüenciació s'usa per dissenyar proteïnes i estudiar mutants (Butterfield et al., 2017; Rocklin et al., 2017; Wrenbeck et al., 2017).

D'altra banda l'avenç aconseguit amb la capacitat i velocitat de computació gràcies a la democratització de la informàtica amb ordinadors personals molt potents i especialment als grans centres de computació com el Barcelona Supercomputing Center-Centre Nacional de Supercomputació (BSC-CNS) situat a la torre Girona de la UPC que correspon a l'antic CESCO. Això permet la representació tridimensional de complexos proteics en tres dimensions de milers d'aminoàcids.

Dins d'aquest context i vista la importància que té l'estudi de les proteïnes en les noves teràpies per malalties rares i tractament innovadors per l'Alzheimer o el càncer vaig decidir endinsar-me en el camp d'estudi de les proteïnes des del punt de vista de la seva cartografia, que és bàsica per assolir els objectius esmentats

1.2 OBJECTIUS DEL TREBALL

El TFM pretén a petita escala i a partir de les dades obtingudes pels mètodes esmentats en l'apartat anterior, i que es troben en forma de taules alfanumèriques emmagatzemats en el Protein Data Bank, representar les proteïnes en el seu format funcional, és a part de la composició química, l'empaquetat tridimensional que li confereix les propietats o funcions que desenvolupen. Una proteïna que ha perdut la seva forma tridimensional es coneix com a proteïna desnaturalitzada que es incapaç de complir la seva funció original.

La lectura per a poder treballar amb els fitxers .pdb es possible fer-la amb la llibreria Biopython però he preferit fer-la mitjançant programació pròpia.

Per tant un cop desgranats tots els tipus de registre de la taula només queda la tasca de representar-la gràficament. Per aquesta tasca s'ha utilitzat el llenguatge de programació Python 2 concretament la versió 2.7.15 i com a llibreries gràfiques la Matplotlib (per a dues dimensions o pseudo 3 dimensions) i la Panda3D (motor de videojocs, per la representació dinàmica en 3D) i en cadascuna de les formes 2D i 3D s'ha programat algunes opcions possibles.

1.3 ENFOCAMENT I MÈTODE SEGUIT

La base del TFM és la representació de l'estructura terciària o quaternària de les proteïnes, és a dir, el seu estat funcional mitjançant la informació subministrada pel PDB (<http://www.rcsb.org/>). La funció del PDB es veu en l'esquema que proposen en la seva pàgina web:

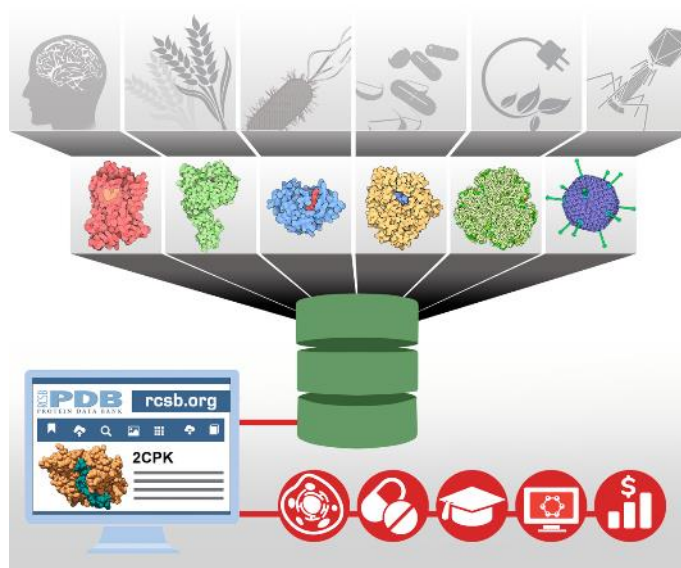


Figura 1. Esquema del flux de treball en l'anàlisi de les proteïnes

L'estructura dels fitxers .pdb és com la que es mostra a continuació:

Pel desenvolupament del projecte s'ha utilitzat l'entorn de desenvolupament integrat IDE utilitzat ha sigut el PyCharm versió 2018.1 com es mostra en la següent imatge:

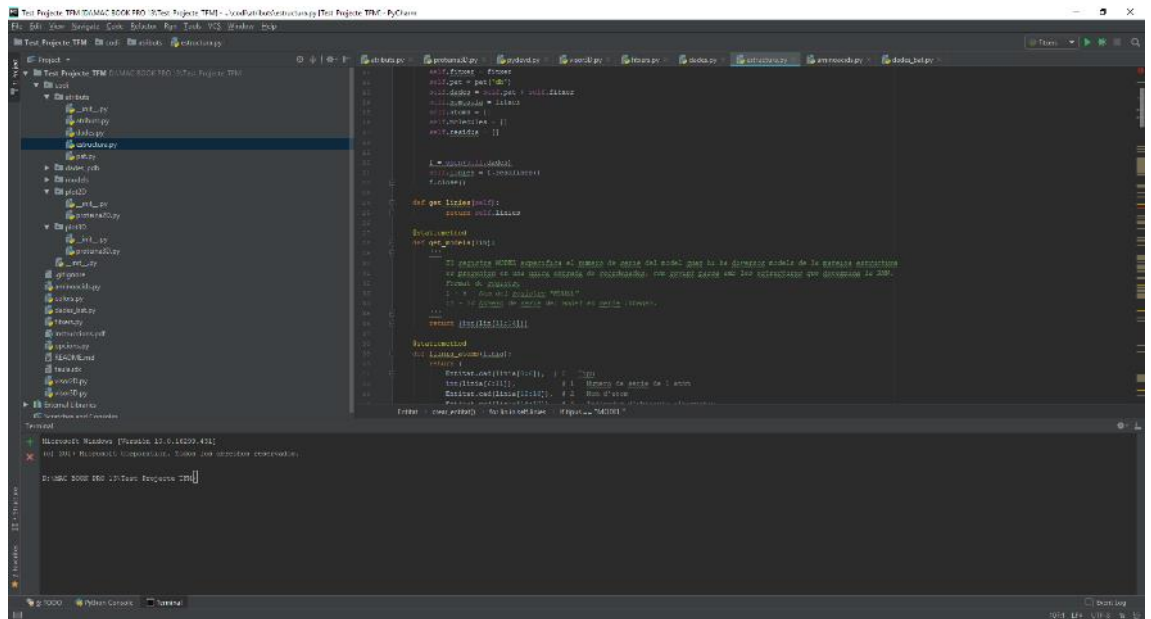


Figura 2 , Entorn IDE PyCharm

Tot i que el programa te moltes opcions i és mol complex, dona moltes facilitats al desenvolupament al poder fer córrer els programes des del mateix editor i utilitzar un debug molt complet, i a l'hora manté una finestra de terminal a sota per treballar directament amb la línia de comandes.

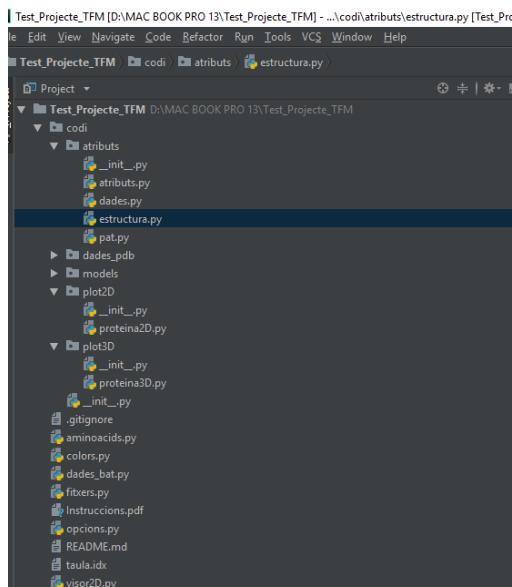


Figura 3. Finestra del projecte amb els mòduls i fitxers auxiliars

El projecte s'ha dividit en els mòduls alfanumèrics i els mòduls gràfics. Entre els mòduls alfanumèrics hi ha els mòduls:

1.3.1 CLASSES ALFANUMÈRIQUES

La classe Entitat que transforma el fitxer .pdb en llistes de python així per exemple hi haurà la llista àtoms, la de residus, de la molècules, la de residus heterogenis, la d'àtoms heterogenis etc. Amb els atributs i mètodes propis de la classe per accedir

Una segona classe anomenada Atributs recull els atributs que corresponen a les llistes i els mètodes per decodificar les llistes obtingudes a la classe Entitat amb els mètodes corresponents

Una tercera classe anomenada Colors , te com atributs la paleta de colors i diferents diccionaris de transformació de atributs a colors de la paleta. Els mètodes són els propis per donar color a cada un dels àtoms en funció de l'atribut escollit

Una quarta classe anomenada Dades, correspondria a la complementaria de la classe Entitat que transforma en llistes alguns tipus de registre que no s'utilitzen en l'entorn gràfic però que serveixen per fer una fitxa de descripció de la proteïna estudiada.

1.3.2 CLASSES GRÀFIQUES

Després hi ha les dues classes gràfiques:

1.3.2.1 CLASSE EN 2D

La classe Plots que utilitza la llibreria Matplotlib per dibuixar la informació obtinguda en la Classe Entitats amb els colors definits en la classe Colors. Existeixen tres mètodes bàsics :

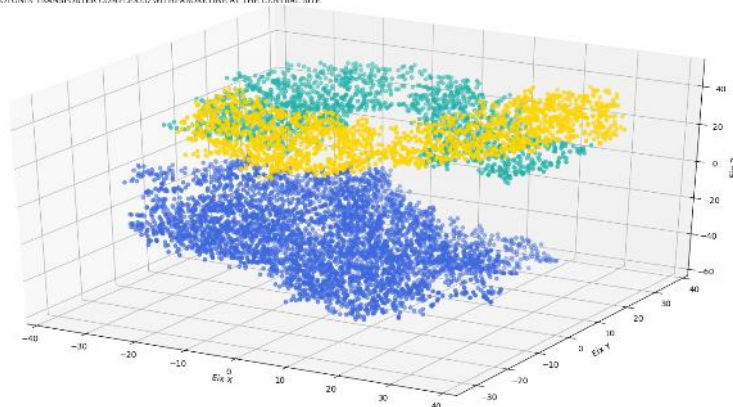
- Àtoms que dibuixa un scatterplot en 3D , ja sia amb els àtoms o amb els residus o i els acolorix segons molècules, tipus d'aminoàcid (hidrofòbics, hidrofílics, àcids o bàsics. Però amb poques línies de codi més es podrien representar els carbonis alfa, determinats tipus d'aminoàcids i acolorir-los amb la paleta de colors de la classe Colors
- Línies. Aquest plot vindria a ser una espècie de representació backbone pels residus de la proteïna amb la senzillesa de presentació que ens dóna el Matplotlib, igual que en el cas anterior existeix la

possibilitat d'acolorir-los segons les característiques dels residus: molècula, tipus, etc.

- Plot2D-3D. Aquest plot es com en el primer cas una representació d'un scatterplot (x,y,z) però amb la particularitat de projectar la visió de l'scatterplot en cadascun dels plans (x,y) (y,z) i (z,x)

Figure 1

X-RAY STRUCTURE OF THE TSH HUMAN SEROTONIN TRANSPORTER COMPLEXED WITH PARoxetine AT THE CENTRAL SITE



31.7682 45.0754 68.1037

Figura 4. Exemple scatterplot de les tres molècules de la proteïna (SERT) 5I6X.pdb

1.3.2.2 CLASSE DE 3D DINÀMICA

Aquesta classe té com a llibreria auxiliar la Panda3D que originàriament va ser creada per Walt Disney. Es tracta d'un motor de videojocs, que es va alliberar sota llicència BSD² l'any 2008.

Aquest és el mòdul més complex de tots ja que a part de la dificultat de dibuixar amb el Panda3D se li sumen les tasques de posicionament de càmera, il·luminació, centrar en els eixos de simetria, i finalment aconseguir els moviments de rotacions en els tres eixos i les funcions apropar i allunyar.

El Panda3D té un problema afegit que només treballa amb imatges planes .png que poden correspondre a textures dels models, els entorns de moviment, paisatges, etc. en 3D que són els fitxers .bam. I els models fixos o animats que estan en un format vectorial .egg. Per construir models com per exemple un cilindre, un con, una hèlix etc. s'ha de dibuixar amb Blender 3D o Maya i després

² és la llicència de programari principalment pels sistemes BSD, pertany al grup de llicències de programari lliure. Aquesta té menys restriccions comparada amb altres com la GPL, sent més propera al domini públic. La llicència BSD, al contrari que la GPL, permet l'ús del codi font en programari privatiu. És molt similar a la llicència MIT. (https://ca.wikipedia.org/wiki/Llic%C3%A8ncia_BSD)

exportar-ho utilitzant un programa inclòs en les tools de panda3D. En aquest punt es troba a faltar que admetés models vectorials més estàndard com el .dxf o el .dgn d'Autocad

La classe utilitzada en aquest mòdul és la classe Proteïna, aquesta classe te dos atributs que fan referencia al moviment el pas de moviment linial i el pas de moviment de rotació dels que depenen la velocitat del moviment ja que son multiplicats per un factor-temps en el mètode de moviment

Els mètodes en general serien els de dibuix, moviment, il·luminació i càmera

1. Mètodes de dibuix **fatomic** que dibuixa els àtoms segons el seu radi (classe Atributs) i els acolorix en funció de l'atribut a representar, un aminoàcid concret, una molècula heterogènia que no forma part de la proteïna, la molècula a la que pertany, etc
2. Mètode **columna** és el mètode que dibuixa els carbonis α ajuntats amb línies. Les esferes que representen els carbonis α són acolorits com en el mètode anterior
3. Mètode **línies** és el mètode pseudo-backbone en el que només existeixen línies que uneixen els diferents carbonis α de cada aminoàcid. El problema per fer un backbone autèntic és el poder dibuixar cilindres i colzes, ho vaig provar amb un cilindre.egg que vaig trobar internet però els atributs per dibuixar-lo eren insuficients per dibuixar angles amb una longitud que no depenges del radi
4. Els mètodes de moviment són el **tecles_accio** i el **moviments** que llegeixen determinades tecles definides del teclat per generar els moviments
5. La il·luminació forma part del cos de la classe i es diferencia amb il·luminació ambiental, il·luminació direccional, il·luminació puntual cadascuna amb les característiques de color posició en l'escena i punt objecte o escena a il·luminar
6. Afegir els eixos de simetria a la proteïna per situar-nos en el moment de gir o translació. Afegir llegendes i títols

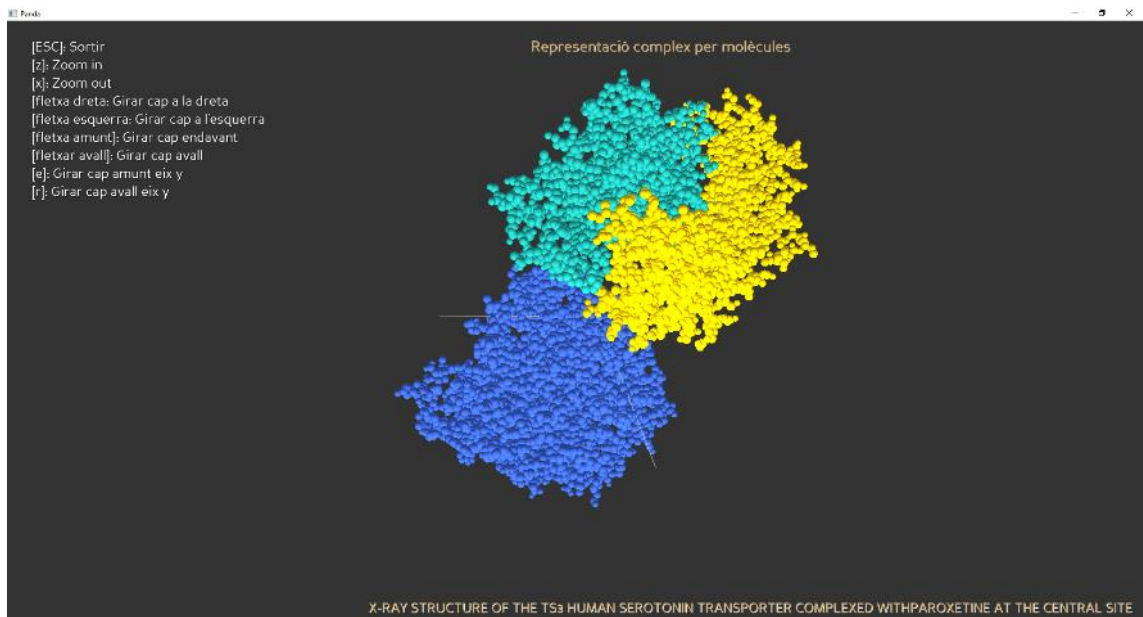


Figura 5. Mateixa proteïna representada anteriorment però visualitzada amb panda3D

1.4 PLANIFICACIÓ DEL TREBALL

La planificació del treball resulta una mica enganyosa perquè el treball de final de màster l'he fet en dos quadrimestres per problemes personals, per tan en el començament del quadrimestre ja tenia bastantes coses fetes o al menys estudiades, i sobretot havia practicat amb python perquè el meu coneixement de phyton es basava en l'assignatura del màster.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
	26-2	5-3	12-3	19-3	26-3	2-4	9-4	16-4	23-4	30-4	7-5	14-5	21-5	28-5	4-6	11-6	18-6	25-6	
	PEC0	PEC1			PEC2			PEC3			PEC4		PEC5a	PEC5b					
Definició projecte																			
Repàs i documentació i soft existent																			
Test software existent																			
Treball amb el fitxer .pdb																			
Escollir plataforma desenvolupament																			
Programació tractar el fitxer .pdb																			
Programació auxiliar atributs																			
Representació amb Matplotlib																			
Representació amb Panda3D																			
Test amb diferents proteïnes																			
Redacció de la memòria																			
Presentació power point																			

1.5 BREU SUMARI DE PRODUCTES OBTINGUTS

El primer producte obtingut correspon a la crida :

```
>>> python dades_bat.py
```

Que crida a la classe Dades i ens dóna una visió general de la proteïna estudiada

```
Taula: 5I6X.pdb .-
Nom complex: X-RAY STRUCTURE OF THE TS3 HUMAN SEROTONIN TRANSPORTER COMPLEXED WITHPAROXETINE AT THE CENTRAL SITE
```

```
-----
Paraules clau: None
```

```
Nombre de models: 1
```

```
Descripció de les molècules:
```

```
-----
Nombre de molècules: ['A', 'B', 'C']
```

```
MOL_ID: 1;
MOLECULE: SODIUM-DEPENDENT SEROTONIN TRANSPORTER;
CHAIN: A;
SYNONYM: SERT,5HT TRANSPORTER,5HTT,SOLUTE CARRIER FAMILY 6 MEMBER 4;
ENGINEERED: YES;
MUTATION: YES;
MOL_ID: 2;
MOLECULE: 8B6 ANTIBODY, HEAVY CHAIN;
CHAIN: B;
ENGINEERED: YES;
MOL_ID: 3;
MOLECULE: 8B6 ANTIBODY, LIGHT CHAIN;
CHAIN: C;
ENGINEERED: YES
```

```
Els elements heterogenis del complex proteic
```

```
-----
Molècula NAG - N-ACETYL-D-GLUCOSAMINE de la molècula A format per 14 àtoms de formula 2(C8 H15 N O6)
Molècula CLR - CHOLESTEROL de la molècula A format per 28 àtoms de formula C27 H46 O
Molècula LMT - DODECYL-BETA-D-MALTOSIDE de la molècula A format per 22 àtoms de formula C24 H46 O11
Molècula 8PR - PAROXETINE de la molècula A format per 24 àtoms de formula C19 H20 F N O3
Molècula CL - CHLORIDE ION de la molècula A format per 1 àtoms de formula CL 1-
Molècula NA - SODIUM ION de la molècula A format per 1 àtoms de formula NA 1+
```

```
ELEMENTS QUIMICS
```

```
-----
Els elements químics que formen part de la proteïna exclosos els hidrògens
```

```
( N : 1202 ) ( C : 4892 ) ( O : 1392 ) ( S : 40 )
Total elements exclosos l'hidrogen: 7526
```

```
Descripció dels elements heterogens:
```

```
-----
Els elements químics que no formen part de la proteïna exclosos l' HOH
( C : 74 ) ( N : 3 ) ( O : 24 ) ( F : 1 ) ( CL : 1 ) ( NA : 1 )
```

```
RESIDUS
```

```
-----
Els residus que formen part de la proteïna que són aminoàcids
```

```
( GLY : 73 ) ( SER : 100 ) ( GLN : 32 ) ( GLU : 35 ) ( ARG : 28 ) ( THR : 85 ) ( TRP : 27 ) ( LYS : 44 ) ( VAL : 71 ) ( ASP : 31 )
( PHE : 57 ) ( LEU : 79 ) ( ILE : 64 ) ( TYR : 53 ) ( ALA : 65 ) ( ASN : 33 ) ( PRO : 47 ) ( CYS : 23 ) ( MET : 17 ) ( HIS : 12 )
Total tipus aminoàcids: 20 i total residus 976
Els residus que formen part de la proteïna que no són aminoàcids
```

```
Els residus heterogenis que no formen part de la proteïna
( NAG : 2 ) ( CLR : 1 ) ( LMT : 1 ) ( 8PR : 1 ) ( CL : 1 ) ( NA : 1 )
```

Process finished with exit code 0

Després la programació d'aquestes opcions de dibuix , la llista és fàcilment ampliable a partir de l'acoloriment segons diferents atributs

Opc	2D i pseudo 3D	Matplotlib	Para 1	Para2
A01.	Representació en pseudo-3d dels a		a	molecula
A02.	Representació en pseudo-3d dels residus (carbonis alfa) per molècules		a	residu
A03.	Representació en 2D i pseudo 3D		a	vistes
A04.	Representació en línies per molècules en pseudo 3D		c	molecula
A05.	Representació en línies dels residus en pseudo 3D segons tipus aminoàcid		c	tipus
>>>	<code>python visor2D.py param1 param 2</code>			

Opció	3D	Panda 3D	Para 1	Param2	Param3
B01.	Representació del complex proteic segons color aminoàcids		a	residu	
B02.	Representació dels elements externs a la proteïna		a	extern	
B03.	Representació per tipus aminoàcids		a	tipus	
B04.	Representació ubicació d'un aminoàcid concret		a	amino	sig_amino
B05.	Representació del complex per molècules		a	molecula	
B06.	Representació dels carbonis alfa units per les línies que representen els		c	alfa	
B07.	Representació en pseudo backbone	c tipus	c	tipus	
>>>	<code>python visor3D.py param1 param 2 [param 3]</code>				

La paleta de colors usada s'obté amb a la classe Colors:

```
>>> python visor3D.py colors.py
```



Figura 6. Paleta de colors

1.6 BREU SUMARI DELS PRODUCTES OBTINGUTS

Aquest apartat crec que en part està contingut en els anteriors els resultats gràfics els deixaré per apartats posteriors.

1.7 BREU DESCRIPCIÓ DELS ALTRES CAPÍTOLS DE LA MEMÒRIA

El capítol 2 recollirà l'estat de l'art de productes que tenen la mateixa finalitat, representar les proteïnes tridimensionalment.

En el capítol 3 es recollirà un resum de l'estructura de la proteïna i funcions

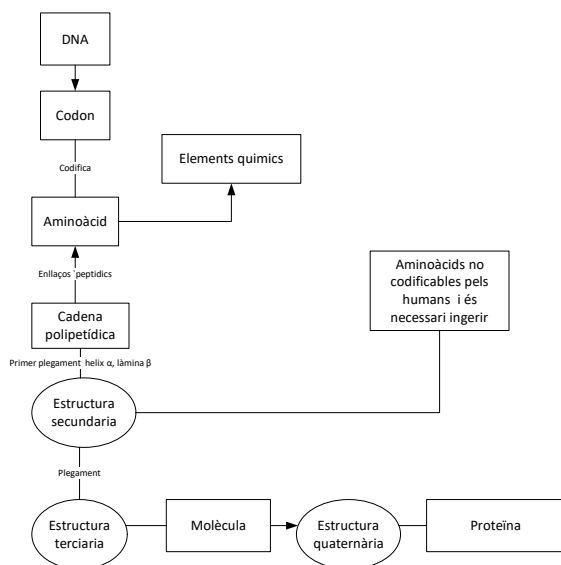
En el capítol 4 es recollirà el treball amb el fitxer del Protein Data Bank amb les seves particularitats

El capítol 5 tractarà de l'ús del Matplotlib com a eina gràfica amb les seves limitacions i els seus avantatges segons la finalitat de l'objecte a representar

Finalment el capítol 6 tractarà de la "relació amor-odi" amb el Panda3D, una eina fantàstica però amb algunes mancances com ja s'han enumerat abans i es detallaran en aquest capítol.

2. ESTRUCTURA DE LES PROTEINES

L'esquema de la Figura 7 es pot veure en els passos de l'estructura de proteïna



L'estructura bàsica o unitat constructiva de la proteïna és l'aminoàcid que adopta un esquema base format per un àtom de carboni (carboni alfa) que forma enllaços amb un grup amino NH_2 , un grup carboxílic COOH amb un doble enllaç entre el segon carboni i un oxigen, un enllaç ocupat per un hidrogen i un grup R que és el que diferencia els 20 aminoàcids diferents.

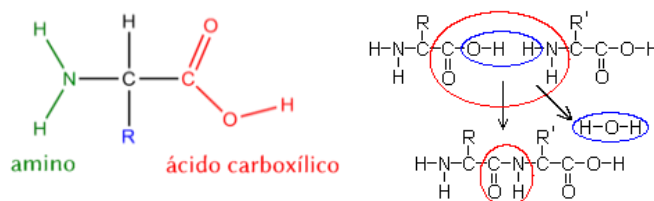


Figura 7. Esquema de la formació de les proteïnes

Figura 8. A la dreta el patró base d'un aminoàcid on es pot veure els dos extrems amb el grup carboxílic i a l'altre extrem el grup amino, essent el braç R que era substituït pels elements que diferencien els diferents aminoàcids. A l'esquerra està la formació de l'enllaç peptídic unint el radical amino d'un aminoàcid amb el carboxílic de l'altra alliberant el molècula d'aigua

Dos aminoàcids reaccionen al ajuntar un grup carboxílic de la primera i el grup amino de la segona alliberant una molècula d'aigua, la llista d'aminoàcids es presenta en el següent quadre:

El resultat d'aquests enllaços peptídics són els 20 aminoàcids següents:

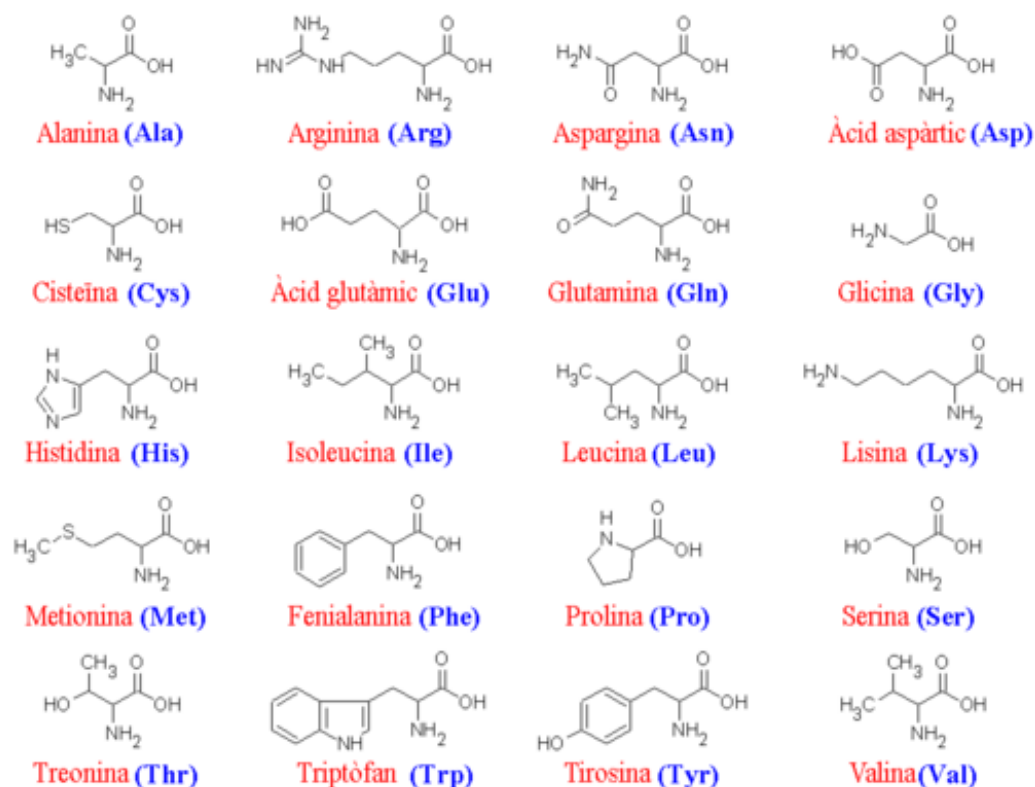


Figura 9.- Els 20 aminoàcids essencials <http://deproteinas.com/los-20-aminoacidos-esenciales>

2.1 ESTRUCTURA PRIMÀRIA DE LA PROTEÏNA

La majoria de les proteïnes un cop sintetitzades no són funcionals. Per ser funcional, tota proteïna ha de plegar-se correctament i, en molts casos, també ha de experimentar modificacions diverses en la seqüència dels seus aminoàcids.

Es diu d'una proteïna plegada correctament que presenta la seva conformació nativa, una diferència d'una altra no plegada o plegada incorrectament, que està en una conformació no nativa. Moltes proteïnes tenen un folding natiu difícil d'aconseguir en un ambient aquós com el de la cèl·lula i, per aconseguir-ho, han d'intervenir les proteïnes xaperones.

La relació entre la seqüència d'aminoàcids que formen una cadena de proteïnes (sovint anomenada estructura primària de la proteïna) això, no obstant, per aconseguir que aquesta cadena es converteixi en una proteïna funcional adoptarà una estructura tridimensional única que constituirà l'estructura nativa de la proteïna.

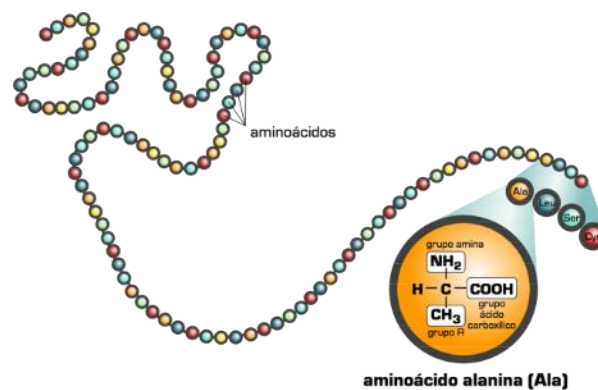


Figura 10.- Estructura primària de la proteïna <http://umaquimicairresistivel.blogspot.com.es/2011/07/nossa-quimica.html>

De la base de dades del PDB és pot obtenir en lloc del format *pdb, es pot obtenir la seqüència FASTA de la proteïna vista fins ara 1TUP

```
>1TUP:F|PDBID|CHAIN|SEQUENCE
ATAATTGGCAAGTCTAGGAA
```

```
>1TUP:E|PDBID|CHAIN|SEQUENCE
TTTCTAGACTTGCCCAATTA
```

```
>1TUP:C|PDBID|CHAIN|SEQUENCE
SSSVPSQKTYQGSYGFRGLFHLHSGTAKSVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGRTRVAMAIYKQSQHMTTEVV
RRCPHHERCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPEPPEVGS DCTTIHYNMNCNSSCMGGMNRRLPILT
IITLEDSSGNLLGRNSFEVRCACPGDRDRRTEENLRKKGEPHHELPPGSTKRALPNNT
```

```
>1TUP:B|PDBID|CHAIN|SEQUENCE
SSSVPSQKTYQGSYGFRGLFHLHSGTAKSVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGRTRVAMAIYKQSQHMTTEVV
RRCPHHERCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPEPPEVGS DCTTIHYNMNCNSSCMGGMNRRLPILT
IITLEDSSGNLLGRNSFEVRCACPGDRDRRTEENLRKKGEPHHELPPGSTKRALPNNT
```

```
>1TUP:A|PDBID|CHAIN|SEQUENCE
SSSVPSQKTYQGSYGFRGLFHLHSGTAKSVTCTYSPALNKMFCQLAKTCPVQLWVDSTPPPGRTRVAMAIYKQSQHMTTEVV
RRCPHHERCSDSDGLAPPQHLIRVEGNLRVEYLDDRNTFRHSVVVPEPPEVGS DCTTIHYNMNCNSSCMGGMNRRLPILT
IITLEDSSGNLLGRNSFEVRCACPGDRDRRTEENLRKKGEPHHELPPGSTKRALPNNT
```

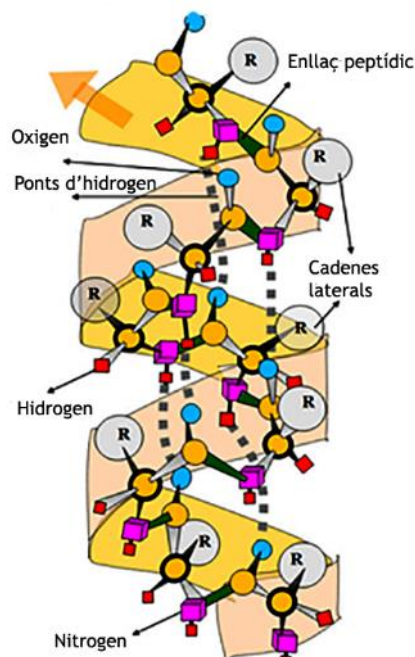
On es pot veure que les dues primeres cadenes 1TUP:F, 1TUP:E corresponen als nucleòtids de la cadena de DNA i la resta a la seqüència

d'aminoàcids que formen la proteïna, les cadenes 1TUP:A,1TUP:B, i 1TUB:C

2.2 ESTRUCTURA SECUNDARIA DE LAS PROTEINAS

Las cadena d'aminoàcids es compacta doblegant-se sobre si mateixa, el primer pas seria el que l'anomenada estructura secundària. Aquest tipus d'estructura bàsicament esta representada per les hèlix α i les làmines β .

La conformació tridimensional és complexa deguda a la dificultat de aquesta relació és la incertesa de la predicció inherent a algunes seqüències d'aminoàcids. És a dir, s'han trobat certes seqüències per adoptar diferents conformacions tridimensionals en diferents condicions. Això implica que les estructures terciàries de les proteïnes, de les quals depenen de les seves funcions, han de ser representades de forma tridimensional a partir de les bases de dades de proteïnes, més concretament. El Protein Data Banc. (<http://www.rcsb.org/>) i més



concretament del format PDB

Figura 11.- Plegament helix α a partir d'enllaços formats per ponts d'hidrogen.
<http://delnutrientealadieta.com/2016/02/03/modulo-5-1-naturaleza-tipos-y-propiedades-nutricionales-de-las-proteinas/19/>

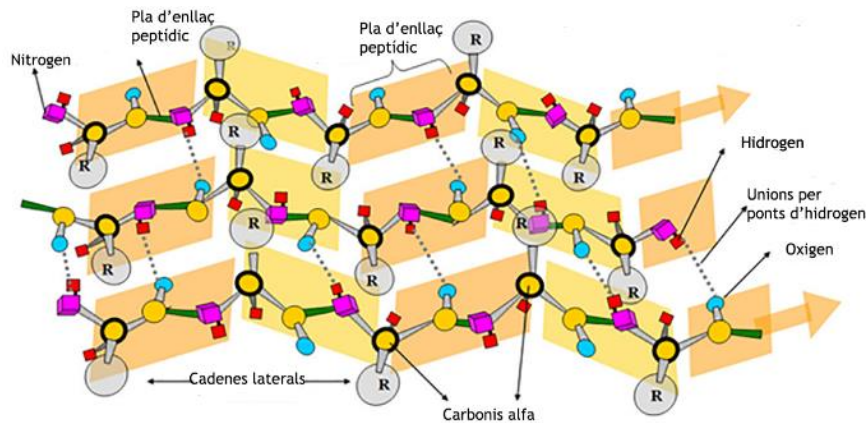


Figura 12 .- Plegament làmines β . <http://delnutrientealadieta.com/2016/02/03/modulo-5-1-naturaleza-tipos-y-propiedades-nutricionales-de-las-proteinas/19/>

L'estructura secundària és el plegament que la cadena polipeptídica adopta gràcies a la formació de ponts d'hidrogen entre els àtoms que formen l'enllaç peptídic. Els ponts d'hidrogen s'estableixen entre els grups (-CO-) i (-NH-) de l'enllaç peptídic (el primer com a acceptor d'H, i el segon com a donador d'H). D'aquesta manera, la cadena polipeptídica és capaç d'adoptar conformacions de menor energia lliure, i per tant, més estables, per tant de l'enllaç peptídic es desprèn una molècula d'aigua

Es poden distingir diversos tipus de conformacions que determinen l'estructura secundària d'una proteïna:

- Hèlix α
- Full β
- Girs β
- Conformació del col·lagen
- Estructures super-secundàries

De les que les dues primeres son les més importants.

Un cop formades les estructures secundaries continua el plegament en les anomenades estructures terciàries mitjançant els enllaços disulfur, interaccions de tipus hidrofòbic, polar , electrostàtica, forces de Van der Waals.

Com a conseqüència de l'establiment d'enllaços peptídics entre els diferents aminoàcids que formen la proteïna s'origina una cadena principal o "esquelet" a partir del qual emergeixen les cadenes laterals dels

aminoàcids (R). Els àtoms que componen la cadena principal de la proteïna són la N (nitrogen) del grup amino (condensat amb l'aminoàcid precedent), la C (carboni) (a partir del qual emergeix la cadena lateral) i el C del grup carboxil (que es condensa amb l'aminoàcid següent). Per tant, la unitat repetitiva bàsica que apareix a la cadena principal d'una proteïna és: (-NH- C-CO-)

2.3 ESTRUCTURA TERCIÀRIA DE LES PROTEÏNES

Es diu estructura terciària a la disposició tridimensional de tots els àtoms que componen la proteïna. L'estructura terciària d'una proteïna és la responsable directa de les seves propietats biològiques, ja que la disposició espacial dels diferents grups funcionals determina la seva interacció amb els diversos lligands.

Per a les proteïnes que consten d'una sola cadena polipeptídica (no tenen estructura quaternària), l'estructura terciària és la màxima informació estructural que es pot obtenir. L'estructura terciària és una disposició precisa i única en l'espai, i sorgeix a mesura que es sintetitza la proteïna. En altres paraules, l'estructura terciària està determinada per la seqüència d'aminoàcids (estructura primària).

Es distingeixen dos tipus d'estructura terciària:

Proteïnes amb estructura terciària de **tipus fibrós** en què una de les dimensions és molt més gran que les altres dues. En són exemples el col·lagen, la queratina del cabell o la fibroïna de la seda, En aquest cas, els elements d'estructura secundària (hèlix α o fulls β) poden mantenir el seu ordenament sense recórrer a grans modificacions, tan sols introduint lleugeres torsions longitudinals, com en els fils d'una corda.

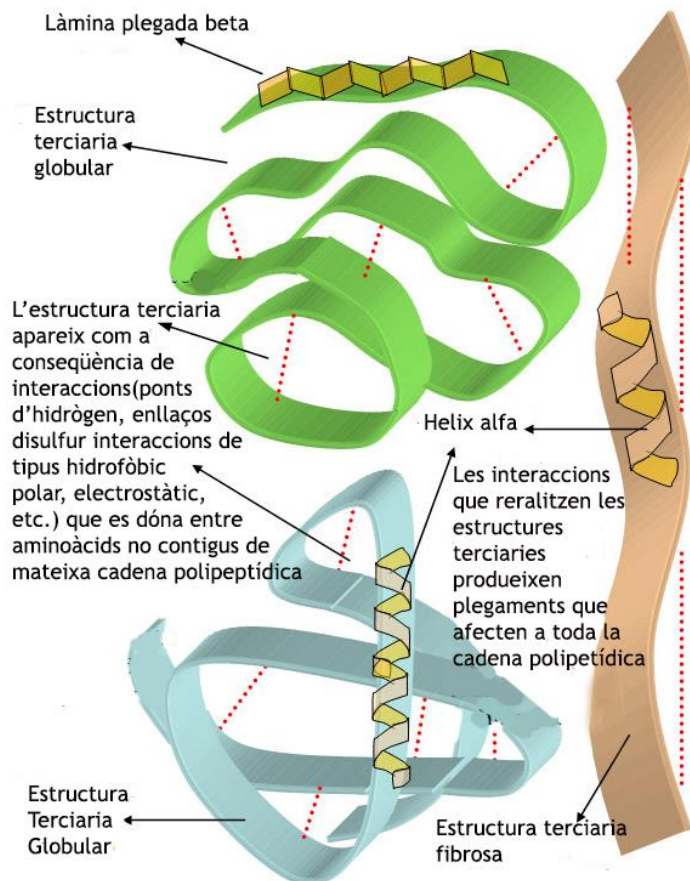


Figura 13.- Plegament terciari. <http://delnutrientealadieta.com/2016/02/03/modulo-5-1-naturaleza-tipos-y-propiedades-nutricionales-de-las-proteinas/19/>

Quan s'uneixen varis pèptids es formen estructures quaternàries com la de l'hemoglobina que es pot veure en la Figura 14

2.4 ESTRUCTURA QUATERNÀRIA DE LES PROTEÏNES

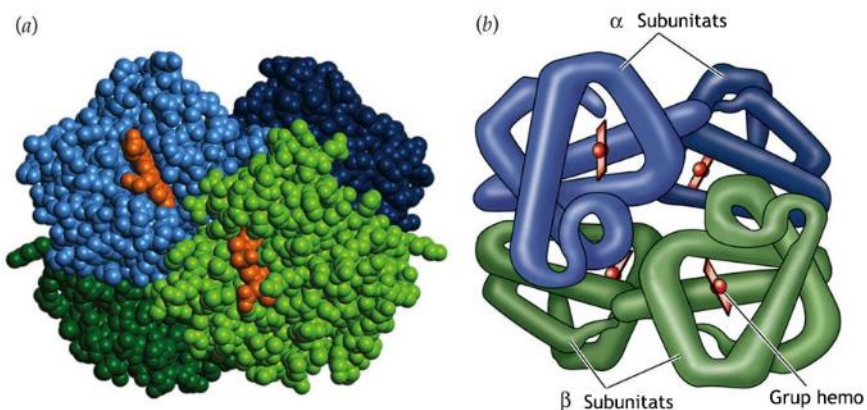


Figura 14.- Estructura quaternària de la proteïna . Exemple de l'hemoglobina. Life: the science of biology , 7^a Edició, 2004

2.5 FUNCIONS DE LES PROTEÏNES

Les funcions de les proteïnes són tan variades com essencials pel funcionament correcte de l'organisme (González, 2018)

- **Funció enzimàtica.** La proteïna actua com a catalitzador de la gran majoria de les reaccions metabòliques específic per a cada reacció. Aquests biocatalitzadors reben el nom d'enzims. La gran majoria de les proteïnes compleixen la funció enzimàtica
- **Funció hormonal.** Algunes hormones són de naturalesa proteica, com la insulina i el glucagó (que regulen els nivells de glucosa en sang) o les hormones segregades per la hipòfisi com l'hormona del creixement, o la calcitonina (que regula el metabolisme del calci).
- **Funció de reconeixement de senyals.**- Les proteïnes un cop secretades exerceixen la seva acció sobre altres cèl·lules dotades d'un receptor adequat.
- **Funció de transport** .-La funció de transport és una funció essencial per portar una molècula hidrofòbica a través d'un mitjà aquós (transport d'oxigen o lípids a través de la sang) o bé per transportar molècules polars a través de barreres hidrofòbiques (transport a través de la membrana plasmàtica).
- **Funció estructural.** Les cèl·lules posseeixen un citoesquelet de naturalesa proteica que constitueix una armadura al voltant del qual s'organitzen tots els seus components, i que dirigeix fenòmens tan importants com el transport intracel·lular o la divisió cel·lular. En els teixits de sosteniment (conjuntiu, ossi, cartilaginós) dels vertebrats, les fibres de col·lagen formen part important de la matriu extracel·lular i són les encarregades de conferir resistència mecànica tant a la tracció com a la compressió.
- **Funció de defensa** La funció fonamental dels mecanismes de defensa és la de discriminar el propi del que és estrany. En bacteris, una sèrie de proteïnes anomenades endonucleases de restricció s'encarreguen d'identificar i destruir aquelles molècules de DNA que no identifica com a pròpies

En els vertebrats superiors, les immunoglobulines s'encarreguen de reconèixer molècules o organismes estranys i s'uneixen a ells per facilitar la seva destrucció per les cèl·lules del sistema immunitari

- **Funció de moviment.**- Les funcions de motilitat dels éssers vius. Així, la contracció del múscul resulta de la interacció entre dues proteïnes, l'actina i la miosina. El moviment de la cèl·lula mitjançant cilis i flagels està relacionat amb les proteïnes que formen els microtúbuls.
- **Funció de reserva.**- La ovoalbúmina de la clara d'ou, la lactoalbúmina de la llet, la gliadina del gra de blat i la hordenina de l'ordi, constitueixen una reserva d'aminoàcids per al futur desenvolupament de l'embrió.
- **Transducció de senyals .**- Els fenòmens de transducció (canvi en la naturalesa fisicoquímica de senyals) estan mediat per proteïnes. Així, durant el procés de la visió, la rodopsina de la retina converteix (o millor dit, transdueix) un fotó lluminós (un senyal física) en un impuls nerviós (un senyal elèctric), i un receptor hormonal converteix un senyal químic (una hormona) en una sèrie de modificacions en l'estat funcional de la cèl·lula.
- **Funció reguladora.**- Moltes proteïnes s'uneixen al DNA i d'aquesta manera controlen la transcripció gènica D'aquesta manera l'organisme s'assegura que la cèl·lula, en tot moment, tingui totes les proteïnes necessàries per exercir normalment les seves funcions. Les diferents fases del cicle cel·lular són el resultat d'un complex mecanisme de regulació exercit per proteïnes com la ciclina.

3. ESTAT DE L'ART

3.1 INTRODUCCIÓ

En https://www.rcsb.org/pages/thirdparty/molecular_graphics hi ha una llista força exhaustiva de 45 visors de molècules la majoria els he testat inclosos els que només són per a Iphone o Ipad

3.2 CC4MG

L'aplicació **CC4mg** té moltes opcions i dóna molta qualitat d'imatge i fins i tot de vídeo i funciona per Windows, MAC OSX(Intel) i Linux. Com a exemple la proteïna 1TUP.pdb

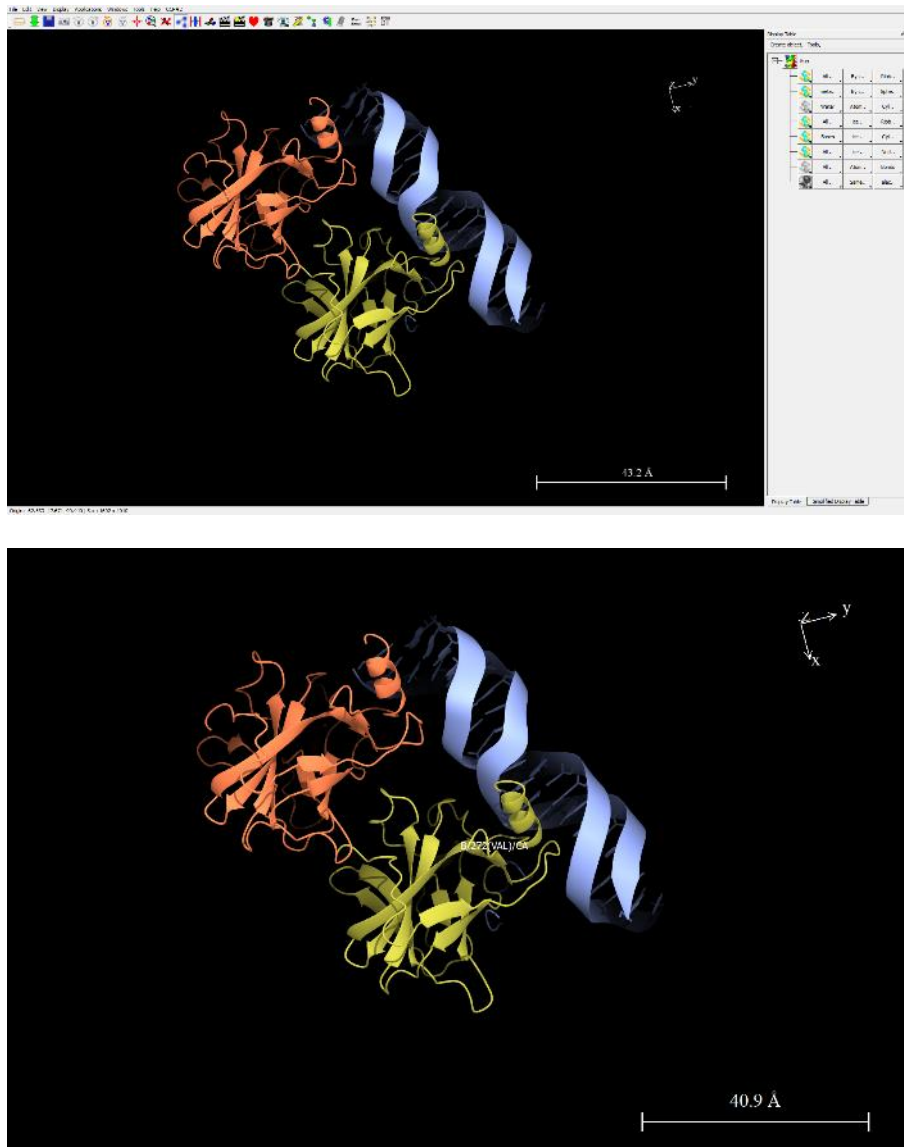


Figura 15- Representació amb CC\$mg supressor de tumors p53 complexat amb ADN

3.4 JMOL I JSMOL

Un programa molt utilitzat és el **Jmol** (java) o **JSmol** (javascript) que generalment no s'utilitzen directament, tot i que té un llenguatge d'scripts per poder augmentar les diferents opcions, generalment s'utilitza a partir amb programes que s'utilitzen a modus d'interfície.

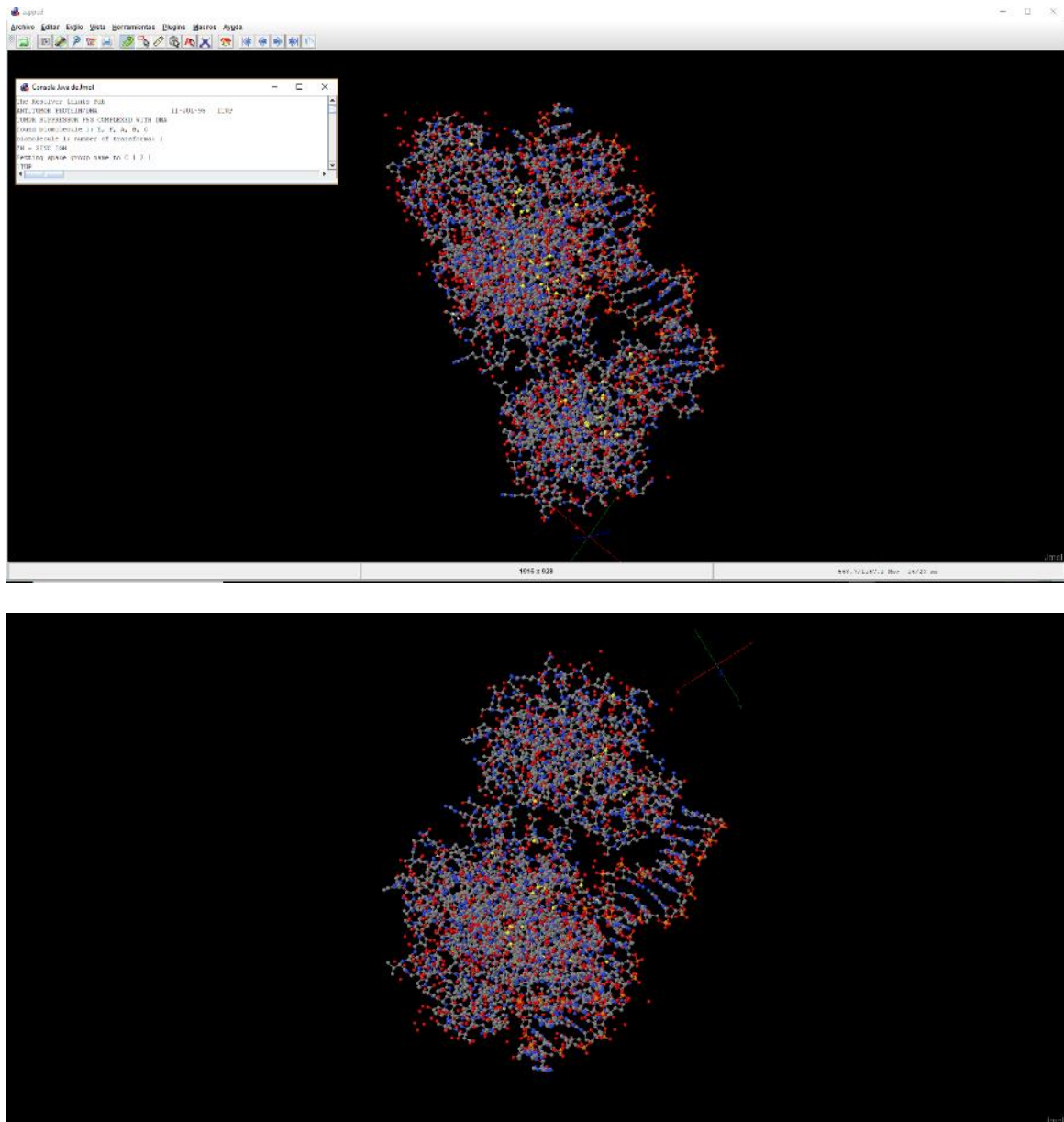


Figura 17- Representació amb Jmol supressor de tumors p53 complexat amb ADN

Una d'aquestes interfícies és el Jena 3D que utilitzar com a visor el JSmol

3.4.1 JENA3D

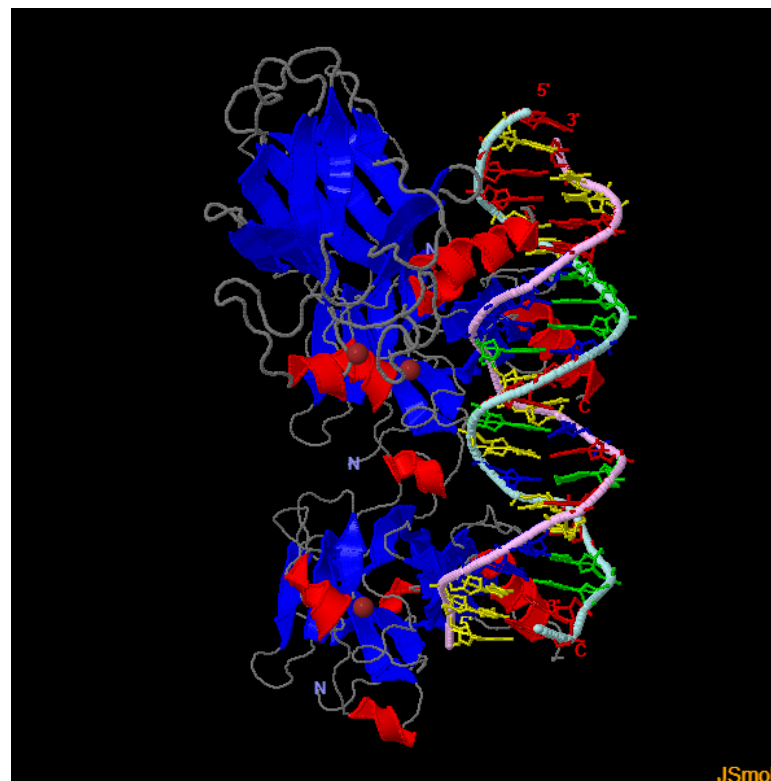
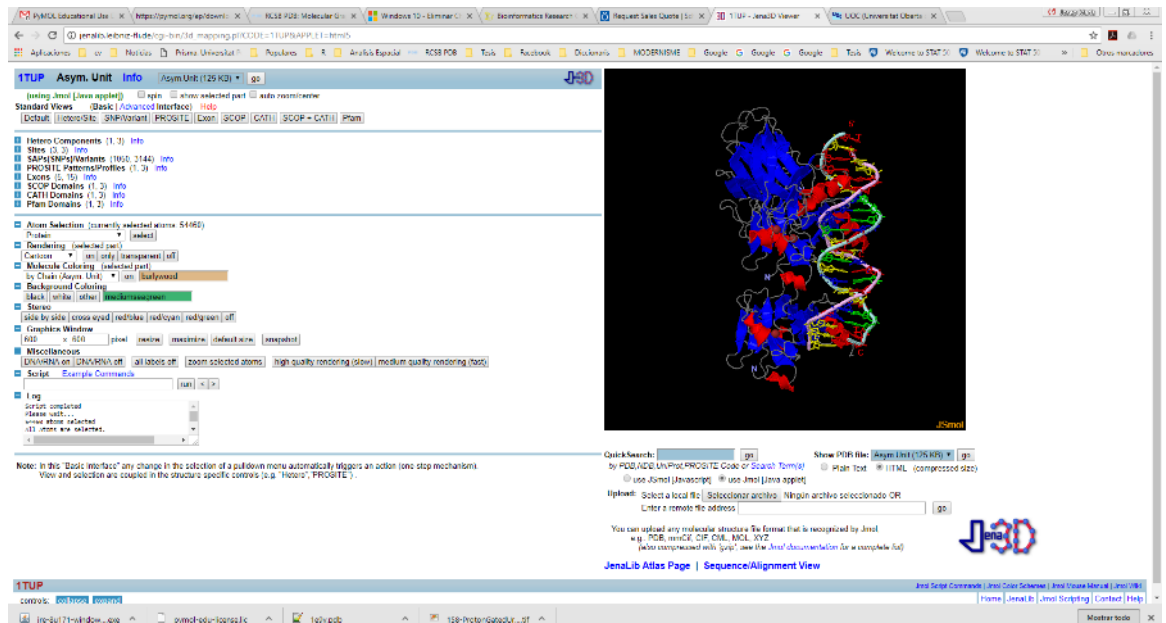
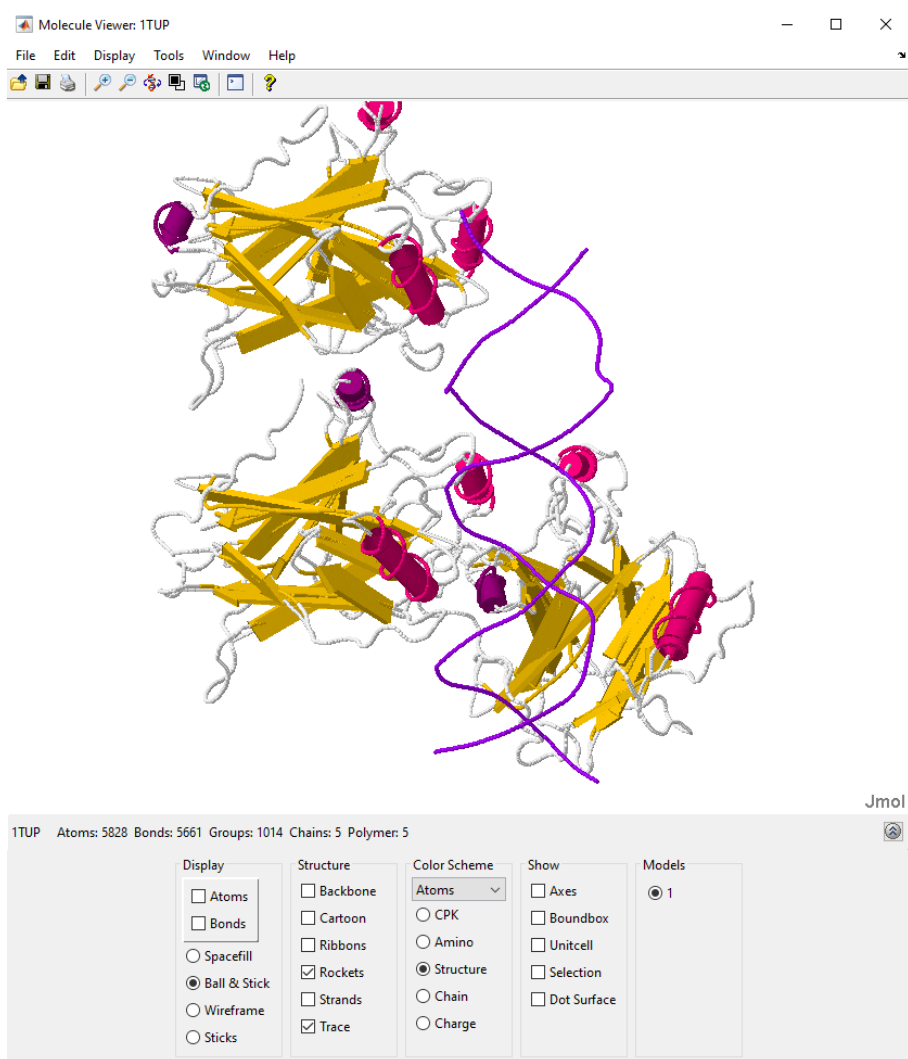
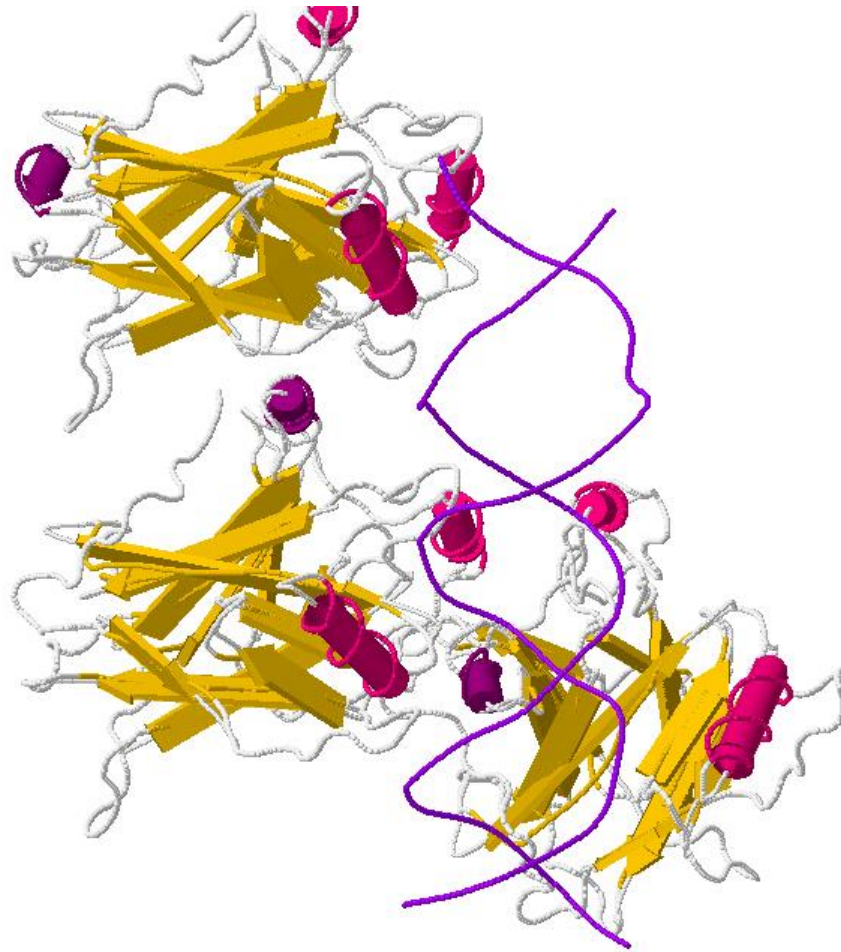


Figura 18- Representació amb Jena3d supressor de tumors p53 complexat amb ADN

3.5 MATLAB





Jmol

Figura 19. Proteïna 1TUP realitzat amb Matlab, com es pot veure el motor Jmol

3.6 EPMV

El programa **ePMV** es un connector de codi obert uPy, el visualitzador molecular de Python incrustat (ePMV), executa programari de modelatge molecular directament en diverses aplicacions professionals d'animació en 3D (hosts). La unió del programari científic i hoste en una sola interfície permet que els usuaris de diversos orígens assemblin visuals de qualitat professional i realitzin experiments computacionals amb relativa facilitat.

La combinació proporciona:

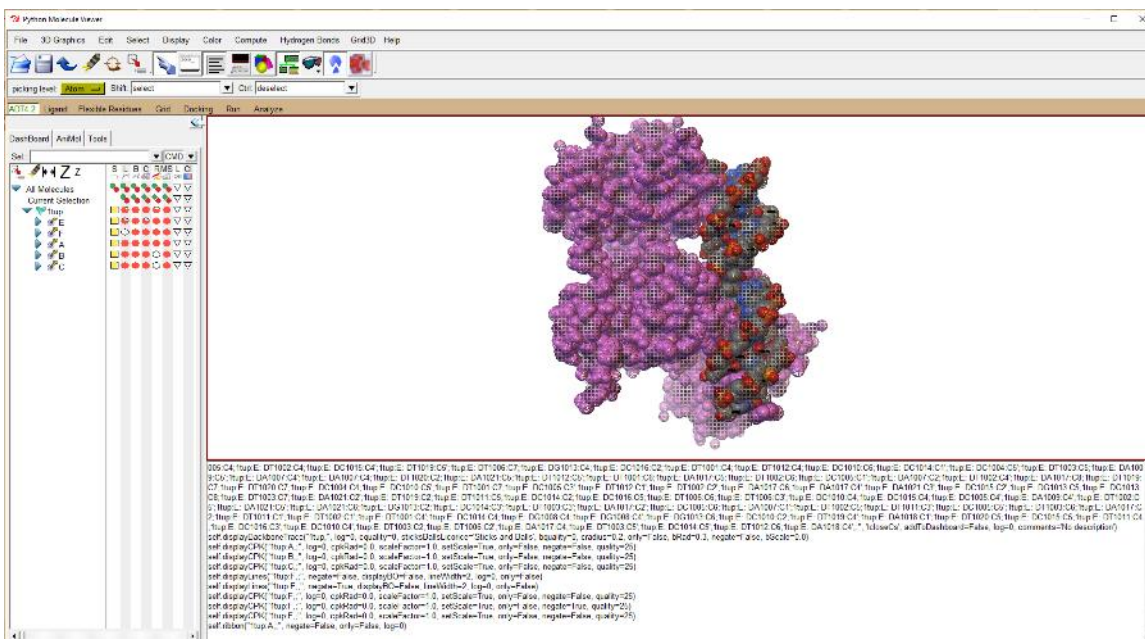
- Representació d'alta qualitat amb ombres, efectes de la càmera, il·luminació global, oclusió ambiental, etc.
- Fluxos de treball intuïtius de la interfície gràfica d'usuari que ajuden als usuaris a configurar animacions que van des de rotacions de placa giratòria fins a sofisticats mecanismes de pel·lícules d'acció

- Modulació de subcarpeta que permet als usuaris il·lustrar o animar esdeveniments cel·lulars complexos en detalls moleculars, posicionant objectes amb controls intuïtius
- Una plataforma Python comú que permet als usuaris iniciar algorismes sofisticats com ara la dinàmica molecular o els càlculs d'energia d'acoblament sobre la marxa i interoperar múltiples algorismes alhora en el mateix model.

- Cinema4D r12, r13, r14 TM
- Blender 2.49, 2.58, 2.60, 2.6.2, 2.6.8a
- Autodesk Maya 2011, 2012, 2013 TM
- Autodesk 3ds Max 2013, 2014 TM

Tot i que no l'he pogut provar per qüestió de temps sembla que pugui ser una opció professional i semi-professional per a realitzar projectes per la seva flexibilitat en l'ús de eines de codi obert sota python. Espero poder-lo comprovar en un futur pròxim després de la pressió del TFM. Perquè a més, com es veurà més endavant, Blender i Autodesk Maya, proporcionen sortides de models utilitzables per una tool de Panda3D per a convertir-los en .egg que és el format utilitzat per panda3D

3.7 PMV (PYTHON MOLECULAR VIEWER)



4. EL PROTEIN DATA BANK I EL TRACTAMENT INFORMÀTIC DELS FITXERS PDB

4.1 TIPUS DE REGISTRES

És possible agrupar registres en categories segons la freqüència amb què apareix el tipus de registre en una entrada. Es tracta del contingut del llibre PDB File Format v. 3.3 pàgines 6-8

4.1.1 UNA INFORMACIO – UNA LINIA

Hi ha registres que només poden aparèixer una vegada i sense continuacions en un fitxer. Ordenats alfabèticament, aquests són:

Tipus registre	Descripció
CRYST1	Paràmetres de cel·la unitària, grup espacial i Z.
END	Últim registre al fitxer.
HEADER	La primera línia de l'entrada conté el codi d'identificació PDB, la classificació i la data de deposició.
NUMMDL	Nombre de models.
MASTER	Registre de control de la comptabilitat.
ORIGXn	Transformació des de coordenades ortogonals fins a les coordenades enviades (n = 1, 2 o 3).
SCALEn	Transformació de coordenades ortogonals a coordenades cristal·logràfiques fraccionals (n = 1, 2 o 3).

És un error que un duplicat d'algun d'aquests registres aparegui en una entrada.

4.1.2 UNA SOLA INFORMACIO, EN UNA O DIVERSES LÍNIES

Són registres que només existeixen conceptualment una vegada en una entrada, però el contingut de la informació pot superar el nombre de columnes disponibles. Per tant, aquests registres continuen en línies posteriors. Ordenats alfabèticament, aquests són:

Tipus registre	Descripció
AUTHOR	Llista de autors de l'estudi
CAVEAT	Indicador d'error de servidor.
COMPND	Descripció dels continguts macromoleculars de l'entrada.
EXPDTA	Tècnica experimental utilitzada per a la determinació de l'estructura.
MDLTYP	Conté anotacions addicionals pertinents a les coordenades presentades a l'entrada.
KEYWDS	Llista de paraules clau que descriuen la macromolècula.
OBSLTE	Declaració que s'ha eliminat l'entrada de la distribució i la llista dels codis d'identificació que la substitueixen.
SOURCE	Font biològica de macromolècules a l'entrada.
SPLIT	Llista d'entrades de PDB que componen complexos macromoleculars més grans.
SPRSDE	Llista d'entrades obsoletes des del llançament públic i reemplaçades per l'entrada actual.
AUTHOR	Llista de contribuents.
CAVEAT	Indicador d'error sever.
COMPND	Descripció dels continguts macromoleculars de l'entrada.
EXPDTA	Tècnica experimental utilitzada per a la determinació de l'estructura.
MDLTYP	Conté anotacions addicionals pertinents a les coordenades presentades a l'entrada.
KEYWDS	Llista de paraules clau que descriuen la macromolècula.
OBSLTE	Declaració que s'ha eliminat l'entrada de la distribució i la llista dels codis d'identificació que la substitueixen.
SOURCE	Font biològica de macromolècules a l'entrada.
SPLIT	Llista d'entrades de PDB que componen complexos macromoleculars més grans.
SPRSDE	Llista d'entrades obsoletes des del llançament públic i reemplaçades per l'entrada actual.
TITLE	Descripció de l'experiment representat a l'entrada.

La segona i les següents línies contenen un camp de continuació, que és un enter just justificat. Aquest nombre s'incrementa en un per cada línia addicional del registre, seguit d'un caràcter en blanc.

4.1.3 VÀRIES VEGADES, UNA LÍNIA

Lla majoria de tipus de registres apareixen diverses vegades, sovint en grups on la informació no està concatenada lògicament, sinó que es presenta en forma de llista. Molts d'aquests tipus de registre tenen un número de sèrie personalitzat que es pot utilitzar, no només per ordenar els registres, sinó també per connectar-se a altres tipus de registre. Ordenats alfabèticament, aquests són:

Tipus registre	Descripció
ANISOU	Factors de temperatura anisotròpics.
ATOM	Registres de coordenades atòmiques per a grups estàndard.
CISPEP	Identificació de residus de pèptids en la conformació cis.
CONNECT	Registres de connectivitat.
DBREF	Referència a l'entrada a la (s) base de dades de la seqüència.
HELIX	Identificació de subestructures helicoidals.
HET	Identificació de grups heterògens no estàndard).
HETATM	Registres de coordenades atòmiques per a heterògens.
LINK	Identificació de lligams inter-residus.
MODRES	Identificació de modificacions als residus estàndard.
MATRIXn	Transformacions que expressen simetria no cristal·logràfica (n = 1, 2 o 3). Hi pot haver diversos conjunts d'aquests registres.
REVDAT	Data de revisió i informació relacionada.
SEQADV	Identificació de conflictes entre PDB i la base de dades de seqüèn anomenada.
SHEET	Identificació de subestructures del full.
SSBOND	Identificació de bons disulfur.

4.1.4 DIVERSES VEGADES, DIVERSES LÍNIES

Hi ha registres que existeixen conceptualment diverses vegades en una entrada, però el contingut d'informació pot excedir el nombre de columnes disponibles. Per tant, aquests registres continuen en línies posteriors. Ordenats alfabèticament, aquests són:

Tipus registre	Descripció
FORMUL	Fórmula química de grups no estàndard.
HETNAM	Nom compost dels heterògens.
HETSYN	Noms compostos sinònims per a heterògens.
SEQRES	Seqüència principal de residus de la columna vertebral.
SITE	Identificació de grups que comprenen importants llocs d'entitat.

La segona i la següent línies contenen un camp de continuació que és un enter just justificat. Aquest nombre s'incrementa en un per cada línia addicional del registre, seguit d'un caràcter en blanc.

Agrupació: hi ha tres tipus de registres utilitzats per agrupar altres registres. Ordenats alfabèticament, aquests són:

Tipus registre	Descripció
ENDMDL	Registre del final del model per a múltiples estructures en una única entrada de coordenades.
MODEL	Especificació del número de model per a estructures múltiples en una única entrada de coordenades.
TER	Terminador de cadenes.

4.1.5 ALTRES:

Els tipus de registre restants tenen una estructura interna detallada. Ordenats alfabèticament, aquests són:

Tipus registre	Descripció
JRNL	Cita de la literatura que defineix el conjunt de coordenades.
REMARK	Observacions generals; poden ser estructurades o de format lliure

4.2 FORMA DE LA TAULA

La taula te diferents formats de registre segons les sis primeres posicions del registre, que és l'únic camp comú en tota la taula, això no obstant hi

ha dos tipus de registres que són idèntics l'ATOM i l'HETATM ja que es tracta en el cas dels àtoms dels elements que formen la proteïna i l'HETATM que correspon als àtoms dels elements heterògens de la proteïna

Seguidament es mostren alguns registres de la taula utilitzats o que podrien ser utilitzats per descriure el complex proteic

Com s'ha pogut veure no hi ha registres propis que corresponguin a aminoàcids, tot i que hi ha la seqüència ordenada dels residus que és el format "SEQRES"

Respecte els elements heterògens que figuren en el complex proteic poden ser diversos des del grup hemo en l'hemoglobina, o un fàrmac com és la paroxetina en el complex 5I6X.pdb "L'estructura de rajos X del transportador de serotonina humana TS3 amb una molècula de paroxetina al lloc central". Un residu heterogeni que s'exclouen les molècules HOH (aigua)

La curiositat d'aquesta proteïna es que forma un complex amb un tros de DNA, que es troba en les dues primeres molècules (chain E i F) i tres molècules proteiques pures

```
HEADER    ANTITUMOR PROTEIN/DNA                11-JUL-95    1TUP
TITLE     TUMOR SUPPRESSOR P53 COMPLEXED WITH DNA
COMPND    MOL_ID: 1;
COMPND    2 MOLECULE: DNA (5' -
COMPND    3 D(*TP*TP*TP*CP*CP*TP*AP*GP*AP*CP*TP*TP*GP*CP*CP*CP*A
COMPND    4 P*AP*TP*TP*A)-3');
COMPND    5 CHAIN: E;
COMPND    6 ENGINEERED: YES;
COMPND    7 MOL_ID: 2;
COMPND    8 MOLECULE: DNA (5' -
COMPND    9 D(*AP*TP*AP*AP*TP*TP*GP*GP*GP*CP*AP*AP*GP*TP*CP*TP*A
COMPND   10 P*GP*GP*AP*A)-3');
COMPND   11 CHAIN: F;
COMPND   12 ENGINEERED: YES;
COMPND   13 MOL_ID: 3;
```

COMPND 14 MOLECULE: PROTEIN (P53 TUMOR SUPPRESSOR);
 COMPND 15 CHAIN: A, B, C;
 COMPND 16 ENGINEERED: YES

KEYWDS ANTIGEN P53, ANTITUMOR PROTEIN/DNA COMPLEX

SEQRES 1 E 21 DT DT DT DC DC DT DA DG DA DC DT DT DG
 SEQRES 2 E 21 DC DC DC DA DA DT DT DA
 SEQRES 1 F 21 DA DT DA DA DT DT DG DG DG DC DA DA DG
 SEQRES 2 F 21 DT DC DT DA DG DG DA DA
 SEQRES 1 A 219 SER SER SER VAL PRO SER GLN LYS THR TYR GLN GLY SER

 SEQRES 16 C 219 LEU ARG LYS LYS GLY GLU PRO HIS HIS GLU LEU PRO PRO
 SEQRES 17 C 219 GLY SER THR LYS ARG ALA LEU PRO ASN ASN THR

HET ZN A 951 1
 HET ZN B 952 1
 HET ZN C 953 1

HETNAM ZN ZINC ION

FORMUL 6 ZN 3(ZN 2+)
 FORMUL 9 HOH *384(H2 O)

HELIX 1 1 SER A 166 HIS A 168 5 3
 HELIX 2 2 PRO A 177 GLU A 180 1 4
 HELIX 3 3 PRO A 278 GLU A 287 1 10
 HELIX 4 4 GLN B 167 THR B 170 5 4
 HELIX 5 5 PRO B 177 ARG B 181 1 5
 HELIX 6 6 PRO B 278 GLU B 287 1 10
 HELIX 7 7 GLY C 105 TYR C 107 5 3
 HELIX 8 8 SER C 166 MET C 169 1 4
 HELIX 9 9 PRO C 177 ARG C 181 1 5
 HELIX 10 10 PRO C 278 GLU C 285 1 8

SHEET 1 A 4 ARG A 110 GLY A 112 0
 SHEET 2 A 4 CYS A 141 TRP A 146 -1 N TRP A 146 O ARG A 110
 SHEET 3 A 4 THR A 230 TYR A 236 -1 N TYR A 234 O CYS A 141
 SHEET 4 A 4 ILE A 195 GLU A 198 -1 N GLU A 198 O HIS A 233

SHEET 1 B 7 CYS A 124 SER A 127 0
 SHEET 2 B 7 LYS A 132 CYS A 135 -1 N PHE A 134 0 THR A 125
 SHEET 3 B 7 LEU A 264 VAL A 274 1 N GLU A 271 0 MET A 133
 SHEET 4 B 7 ILE A 251 GLU A 258 -1 N LEU A 257 0 LEU A 265
 SHEET 5 B 7 ARG A 156 TYR A 163 -1 N ILE A 162 0 LEU A 252
 SHEET 6 B 7 HIS A 214 PRO A 219 -1 N VAL A 218 0 VAL A 157
 SHEET 7 B 7 GLU A 204 ASP A 207 -1 N LEU A 206 0 SER A 215
 SHEET 1 F 7 THR C 125 SER C 127 0
 SHEET 2 F 7 LYS C 132 GLN C 136 -1 N PHE C 134 0 THR C 125
 SHEET 3 F 7 LEU C 264 CYS C 275 1 N GLU C 271 0 MET C 133
 SHEET 4 F 7 ILE C 251 GLU C 258 -1 N LEU C 257 0 LEU C 265
 SHEET 5 F 7 ARG C 156 TYR C 163 -1 N ILE C 162 0 LEU C 252
 SHEET 6 F 7 HIS C 214 PRO C 219 -1 N VAL C 218 0 VAL C 157
 SHEET 7 F 7 GLU C 204 ASP C 207 -1 N LEU C 206 0 SER C 215

ATOM 1 O5' DT E1001 25.930 6.224 81.661 1.00 67.33 O
 ATOM 2 C5' DT E1001 25.549 7.533 82.107 1.00 69.16 C
 ATOM 3 C4' DT E1001 25.637 8.573 80.994 1.00 65.63 C
 ATOM 4 O4' DT E1001 24.711 9.640 81.304 1.00 69.89 O
 ATOM 5 C3' DT E1001 26.967 9.318 80.881 1.00 69.57 C
 ATOM 6 O3' DT E1001 27.888 8.691 79.993 1.00 68.64 O
 ATOM 7 C2' DT E1001 26.577 10.716 80.386 1.00 65.62 C
 ATOM 8 C1' DT E1001 25.059 10.737 80.490 1.00 68.50 C

.....
 ATOM 419 N3 DA E1021 80.741 44.131 82.312 1.00 55.18 N
 ATOM 420 C4 DA E1021 80.338 44.977 83.265 1.00 56.41 C
 TER 421 DA E1021

ATOM 422 O5' DA F1101 79.715 51.383 69.400 1.00 67.58 O
 ATOM 423 C5' DA F1101 79.560 51.854 70.749 1.00 70.00 C

.....
 ATOM 705 N3 DT F1114 43.616 28.242 81.686 1.00 67.33 N
 ATOM 706 C4 DT F1114 43.224 28.678 80.424 1.00 64.57 C
 ATOM 707 O4 DT F1114 42.455 29.630 80.319 1.00 62.75 O
 TER 857 DA F1121

.....
 ATOM 858 N SER A 94 76.857 22.015 81.259 1.00 23.50 N
 ATOM 859 CA SER A 94 75.562 21.797 80.653 1.00 24.59 C
 ATOM 860 C SER A 94 75.707 20.934 79.426 1.00 26.11 C

```

ATOM   861  O   SER A  94      76.822  20.663  78.986  1.00  23.93      O
ATOM   862  CB  SER A  94      74.924  23.129  80.295  1.00  22.31      C
ATOM   863  OG  SER A  94      74.553  23.776  81.497  1.00  20.52      O
ATOM   876  N   VAL A  97      76.750  21.858  73.095  1.00  22.57      N
ATOM   877  CA  VAL A  97      76.057  22.298  71.922  1.00  23.40      C
ATOM   878  C   VAL A  97      77.101  22.766  70.944  1.00  27.15      C
ATOM   879  O   VAL A  97      78.216  22.994  71.372  1.00  29.57      O
ATOM   880  CB  VAL A  97      75.303  23.518  72.378  1.00  29.18      C
ATOM   881  CG1 VAL A  97      73.974  23.106  72.990  1.00  31.07      C
ATOM   882  CG2 VAL A  97      76.178  24.181  73.431  1.00  34.23      C
.....
ATOM   5438 N   LEU C  289      22.393  11.002  97.959  1.00  70.00      N
ATOM   5439 CA  LEU C  289      22.561  10.184  99.137  1.00  68.50      C
ATOM   5440 C   LEU C  289      22.118  11.007 100.337  1.00  70.00      C
ATOM   5441 O   LEU C  289      23.001  11.323 101.176  1.00  70.00      O
ATOM   5442 CB  LEU C  289      24.076   9.896  99.270  1.00  69.65      C
ATOM   5443 CG  LEU C  289      24.538   8.437  99.265  1.00  68.61      C
ATOM   5444 CD1 LEU C  289      23.800   7.615  98.214  1.00  70.00      C
ATOM   5445 CD2 LEU C  289      26.048   8.378  99.026  1.00  68.78      C
TER    5446      LEU C  289
.....
HETATM 5447 ZN    ZN A  951      58.108  23.242  57.424  1.00  27.10      ZN
HETATM 5448 ZN    ZN B  952      60.108  17.981  75.931  1.00  27.98      ZN
HETATM 5449 ZN    ZN C  953      33.653   0.403  74.115  1.00  41.26      ZN
HETATM 5450 O   HOH E2024      62.701  30.637  78.324  1.00  24.17      O
.....
MASTER      516   0   3  10  33   0   3  15 5828   5  15  55
END

```

4.2.1 LA FITXA GENERAL QUE TREU L'APLICACIÓ

La fitxa general que treu de la taula és

ANTITUMOR PROTEIN/DNA 11-JUL-95 1TUP

Taula: 1tup.pdb .- Nom complex: TUMOR SUPPRESSOR P53 COMPLEXED WITH DNA

Paraules clau: ANTIGEN P53, ANTITUMOR PROTEIN/DNA COMPLEX

Nombre de models: 1

Descripció de les molècules:

Nombre de molècules: ['E', 'F', 'A', 'B', 'C']

MOL_ID: 1;
MOLECULE: DNA (5'-
D(*TP*TP*TP*CP*CP*TP*AP*GP*AP*CP*TP*TP*GP*CP*CP*CP*A
P*AP*TP*TP*A)-3');
CHAIN: E;
ENGINEERED: YES;
MOL_ID: 2;
MOLECULE: DNA (5'-
D(*AP*TP*AP*AP*TP*TP*GP*GP*GP*CP*AP*AP*GP*TP*CP*TP*A
P*GP*GP*AP*A)-3');
CHAIN: F;
ENGINEERED: YES;
MOL_ID: 3;
MOLECULE: PROTEIN (P53 TUMOR SUPPRESSOR);
CHAIN: A, B, C;
ENGINEERED: YES

Els elements heterògens del complex proteic

Molècula ZN - ZINC ION de la molècula A format per 1 àtoms de
formula 3(ZN 2+)

ESTRUCTURES SECUNDÀRIES

Hi ha 10 hèlix alfa
Hi ha 6 làmines beta i 33 subestructures de lamina

ELEMENTS QUÍMICS

Els elements químics que formen part de la proteïna exclosos els
hidrògens

(O : 1114) (C : 3238) (N : 1001) (P : 40) (S : 48)
Total elements exclosos l'hidrogen: 5441

Descripció dels elements heterògens:

Els elements químics que no formen part de la proteïna exclosos l' HOH
(ZN : 3)

RESIDUS

Els residus que formen part de la proteïna que són aminoàcids

(SER : 54) (VAL : 45) (PRO : 42) (GLN : 21) (LYS : 15) (THR
: 42) (TYR : 24) (GLY : 39) (PHE : 15) (ARG : 51)

(LEU : 42) (HIS : 21) (ALA : 21) (CYS : 30) (ASN : 27) (MET : 18) (TRP : 3) (ASP : 24) (ILE : 18) (GLU : 33)
Total tipus aminoàcids: 20 i total residus 585

Els residus que formen part de la proteïna que no són aminoàcids
(DT : 13) (DC : 8) (DA : 13) (DG : 8)

Els residus heterogenis que no formen part de la proteïna
(ZN : 3)

4.3 ESTRUCTURA DE LES DADES

L'estructura de les dades es pot veure en la Figura 20, la proteïna pot estar composta per 1 o més models. Cada model està formada per un conjunt

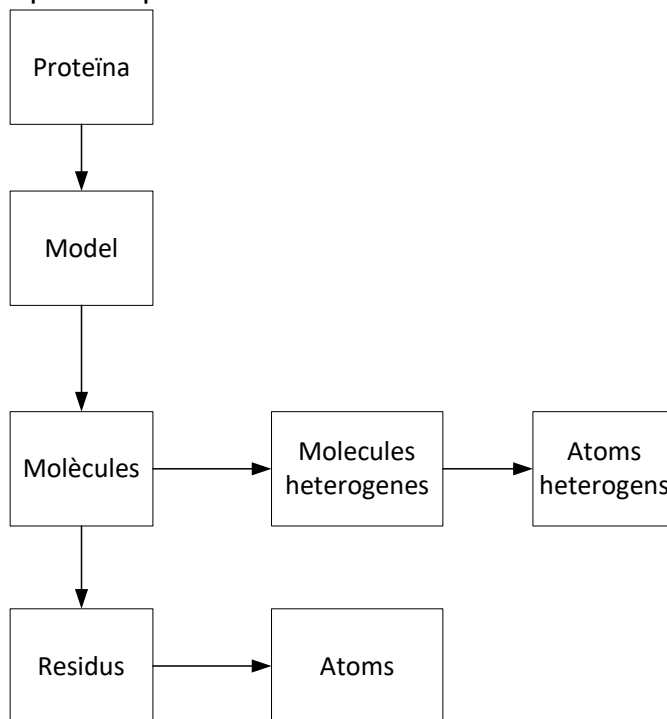


Figura 20.- Estructura de les dades

de molècules (que es la seva estructura quaternària). Cada molècula està composta per un conjunt d'aminoàcids o altres elements com els àcids nucleics .

Cada aminoàcid està compost de varis àtoms que corresponen al format registre "ATOM ". També s'inclouen els elements heterògens que corresponen al registre encapçalada "HETATM"

El format de registres dels registres "ATOM" i "HETATM"

El format del registre és

Columnnes	Tipus de dada	Camp	Descripció
0:6	Nom registre	"ATOM " o "HETATH"	
6:11	Enter	serial	Número de sèrie de l'àtom.
12:16	Atom	name	Nom d'àtom
16:17	Caràcter	altLoc	Indicador d'ubicació alternatiu.
17:20	Nom residu	resName	Nom de residu.
21:22	Caràcter	chainID	Identificador de la molècula.
22:26	Enter	resSeq	Número de seqüència de residus.
26:27	Un caràcter	iCode	Codi d'inserció de residus.
30:38	Num Real(8,3)	x	Coordenades ortogonals per X en Àngstroms.
38:46	Num Real(8,3)	y	Coordenades ortogonals per a Y en Àngstroms.
46:54	Num Real(8,3)	z	Coordenades ortogonals per Z en Àngstroms.
54:60	Num Real(6,2)	occupancy	Ocupació.
60:66	Num Real(6,2)	tempFacto	Factor de temperatura.
76:78	Cadena 2 carac	element	Símbol d'element, justificat.
78:80	Cadena 2 carac	charge	Càrrega a l'àtom.

Com exemple agafarem els quatre primers residus de la molècula A:

```
SEQRES 1 A 219 SER SER SER VAL PRO SER GLN LYS THR TYR GLN GLY SER
```

que es representa en els següents registres del tipus "ATOM", una característica molt important és que els registres seqüencials estan ordenats, és a dir no s'ha de ordenar per cap camp.

Els quatre primers residus veiem en el registre de "SEQRES" són tres serines de fórmula $C_3H_7NO_3$ i una valina $C_5H_{11}NO_2$, com es podrà veure en la taula falten els hidrògens i un oxigen per que ja s'ha format l'enllaç peptídic,

Tipus	serial	name	altLoc	resName	chainID	resSeq	iCode	x	y	z	occupa	temp Factor	element	charge
ATOM	858	N		SER	A	94		76.857	22.015	81.259	1.00	23.50	N	
ATOM	859	CA		SER	A	94		75.562	21.797	80.653	1.00	24.59	C	
ATOM	860	C		SER	A	94		75.707	20.934	79.426	1.00	26.11	C	
ATOM	861	O		SER	A	94		76.822	20.663	78.986	1.00	23.93	O	
ATOM	862	CB		SER	A	94		74.924	23.129	80.295	1.00	22.31	C	
ATOM	863	OG		SER	A	94		74.553	23.776	81.497	1.00	20.52	O	
ATOM	864	N		SER	A	95		74.589	20.502	78.867	1.00	29.04	N	
ATOM	865	CA		SER	A	95		74.668	19.703	77.670	1.00	23.53	C	
ATOM	866	C		SER	A	95		75.458	20.479	76.621	1.00	23.17	C	
ATOM	867	O		SER	A	95		75.180	21.647	76.372	1.00	23.61	O	
ATOM	868	CB		SER	A	95		73.279	19.381	77.166	1.00	21.26	C	

Tipus	serial	name	altLoc	resName	chainID	resSeq	iCode	x	y	z	occupa	temp Factor	element	charge
ATOM	869	OG		SER	A	95		73.128	17.977	77.161	1.00	31.17	O	
ATOM	870	N		SER	A	96		76.466	19.855	76.028	1.00	24.75	N	
ATOM	871	CA		SER	A	96		77.257	20.526	75.012	1.00	26.68	C	
ATOM	872	C		SER	A	96		76.406	20.794	73.792	1.00	29.59	C	
ATOM	873	O		SER	A	96		75.476	20.051	73.494	1.00	36.22	O	
ATOM	874	CB		SER	A	96		78.472	19.725	74.605	1.00	31.31	C	
ATOM	875	OG		SER	A	96		78.231	18.364	74.880	1.00	36.85	O	
ATOM	876	N		VAL	A	97		76.750	21.858	73.095	1.00	22.57	N	
ATOM	877	CA		VAL	A	97		76.057	22.298	71.922	1.00	23.40	C	
ATOM	878	C		VAL	A	97		77.101	22.766	70.944	1.00	27.15	C	
ATOM	879	O		VAL	A	97		78.216	22.994	71.372	1.00	29.57	O	
ATOM	880	CB		VAL	A	97		75.303	23.518	72.378	1.00	29.18	C	
ATOM	881	CG1		VAL	A	97		73.974	23.106	72.990	1.00	31.07	C	
ATOM	882	CG2		VAL	A	97		76.178	24.181	73.431	1.00	34.23	C	

De la taula s'han de destacar el camp **resseq**, que és el que defineix el residu, és pot veure que es tracta dels residus 94, 95, 96 i 97. L'altre camp important és el camp **name** que és el codi de l'element químic més si està més d'una vegada se li afegeix un caràcter més. En el cas dels carbonis de la Valina hi ha el **CA**, que és el carboni α , els altres s'anomenen **CG1**, **CG2** perquè estan numerats ja que es tracta de dos metils CH_3 i per tant es diferencien amb el número final.

Els altres camps importants són, evidentment els camps **x**, **y** i **z** que corresponen a les coordenades de l'àtom, el **resName** que son les sigles

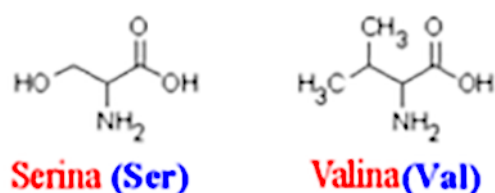


Figura 21. Fòrmules dels aminoàcids Serina i Valina

del residu i el camp **element**, ja que es pot tractar independentment i per la representació necessitarem el radio de l'àtom i el **chainID** que generalment es representa amb una lletra majúscula.

Els altres camps importants són, evidentment els camps **x**, **y** i **z** que corresponen a les coordenades de l'àtom, el **resName** que son les sigles del residu i el camp **element**, ja que es pot tractar independentment i per

la representació necessitem el radio de l'àtom i el **chainID** que generalment es representa amb una lletra majúscula.

Per tant un residu estarà format pels àtoms que tenen el mateix **chainID** i el **resseq** i les seves coordenades correspondran, per decisió meua les del carboni alfa name == CA, això no obstant es podria agafar la mitjana de les coordenades de tots els àtoms, així en el cas del triptòfan que és l'aminoàcid amb més massa alta que té la fórmula:

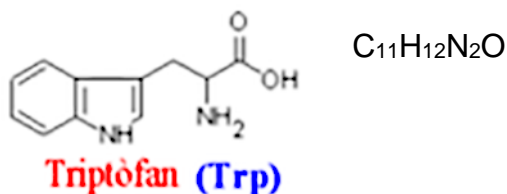


Figura 22.- Fòrmula del Triptòfan

Per validar la hipòtesis s'agafa un triptòfan de la proteïna 1TUP, s'obté la mitjana de les coordenades i la diferència entre la mitjana de les coordenades totals i la de dels carboni alfa són d'1 Armstrong en les coordenades x i z i 2

Armstrongs en la coordenada y

Tipus	serial	name	altLoc	resName	chainID	resSeq	iCode	x	y	z	occupa	tempFacto	element	charge
ATOM	1250	N		TRP	A	146		84,733	35,587	52,259	1	30,23	N	
ATOM	1251	CA		TRP	A	146		85,816	36,553	52,162	1	30,3	C	
ATOM	1252	C		TRP	A	146		86,998	35,943	52,888	1	28,42	C	
ATOM	1253	O		TRP	A	146		86,865	35,433	54,001	1	28,31	O	
ATOM	1254	CB		TRP	A	146		85,45	37,952	52,718	1	37,69	C	
ATOM	1255	CG		TRP	A	146		84,442	38,584	51,817	1	52,88	C	
ATOM	1256	CD1		TRP	A	146		83,104	38,352	51,823	1	59,08	C	
ATOM	1257	CD2		TRP	A	146		84,69	39,461	50,711	1	56,67	C	
ATOM	1258	NE1		TRP	A	146		82,495	39,051	50,813	1	60,65	N	
ATOM	1259	CE2		TRP	A	146		83,448	39,737	50,111	1	58,53	C	
ATOM	1260	CE3		TRP	A	146		85,839	40,041	50,19	1	57,95	C	
ATOM	1261	CZ2		TRP	A	146		83,337	40,587	49,016	1	60,7	C	
ATOM	1262	CZ3		TRP	A	146		85,733	40,876	49,103	1	59,69	C	
ATOM	1263	CH2		TRP	A	146		84,494	41,148	48,524	1	61,66	C	
								Mitjana	84,817	38,522	51,153			
								Diferencia amb CA	0,999	-1,969	1,009			

Com es pot veure en la figura anterior en les representacions de les proteïnes s'agafa el CA perquè els radicals s'allunyen de l'esquelet de la proteïna

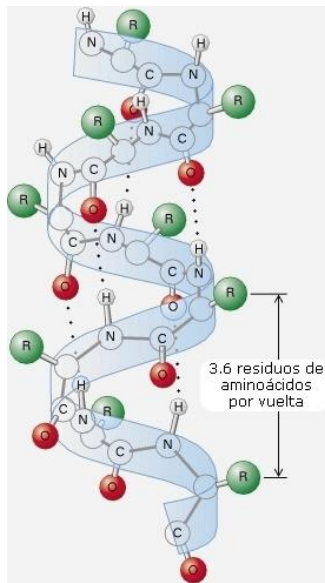
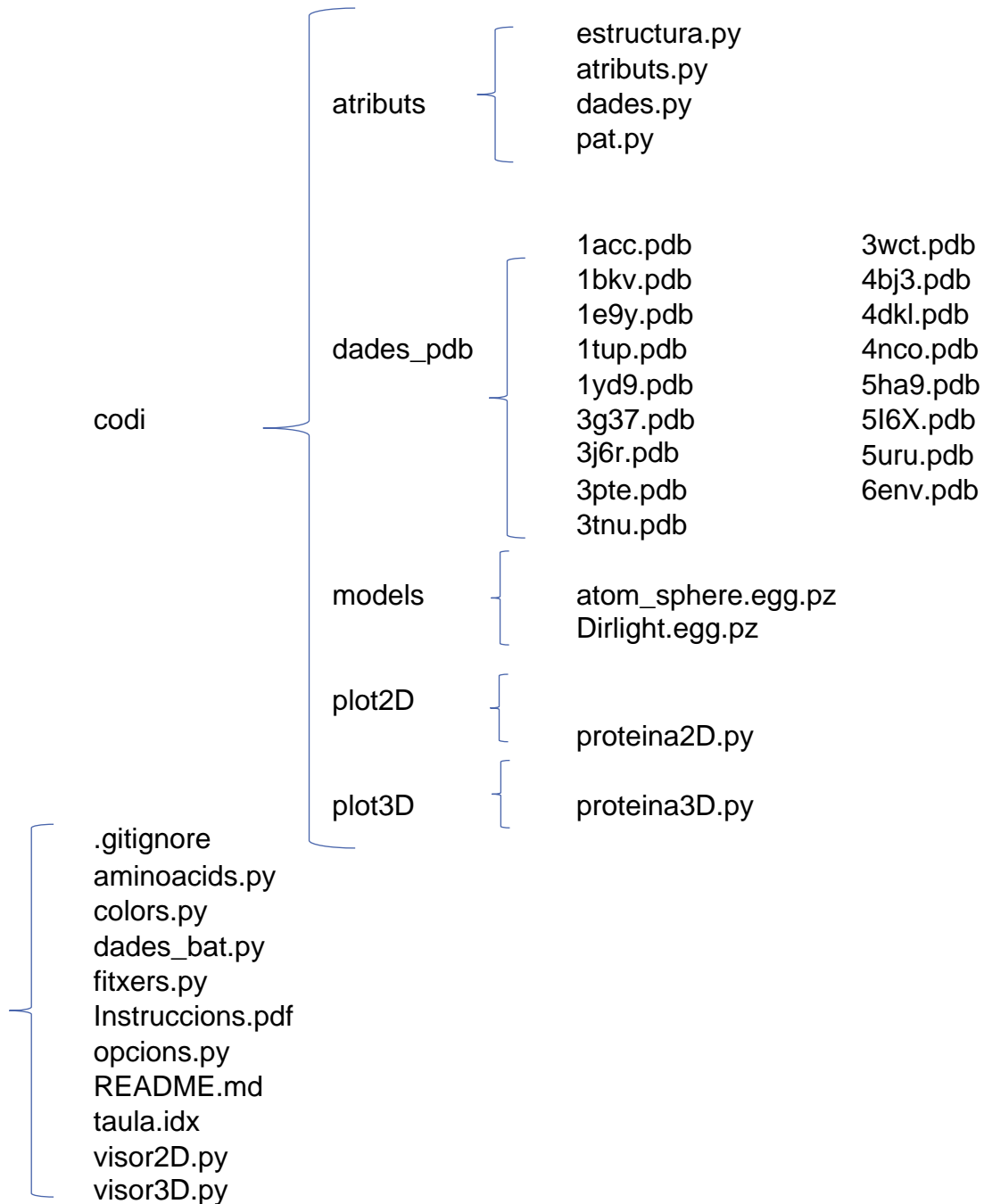


Figura 23.- Tros d'estructura secundaria. <https://maestradelia.files.wordpress.com/2012/11/ch3f6.jpg>

5. ESTRUCTURA DEL PROGRAMA

L'estructura del programa és la següent:



El conjunt del programa es pot dividir en dues parts els mòduls principals que són els que es troben dins de codi juntament amb les dades d'exemple i els models de dibuix 3D, la segona part és la de aplicacions d'ajuda o de crida a les contingudes de codi.

5.1 CODI.ESTRUCTURA

5.1.1 CLASSE ENTITAT

La classe entitat rep el nom de la taula

5.1.1.1 ATRIBUTS

Cap

5.1.1.2 METODEDES

1. **get_linies()**

Entrada: el nom de la taula

Retorna: una llista amb tots els registres de corresponent a la taula

2. **get_models()**

Entrada: registre de la taula si el tipus de registre és "MODEL "

Sortida: número de models (si no en troba cap retorna 1)

Entrada: el nom de la taula

Retorna: una llista amb tots els registres de corresponent a la taula

3. **flinia_atoms()**

Entrada: línia del registre llegit de la taula

Retorna: una llista de camps que representen el registre, el nom dels camps està en el mètode `camps_atoms` de la classe Atributs

4. **cad()**

Entrada: cadena

Retorna: cadena neta de espais al començament i al final i dels codis d'escape `\t`, `\n`, `\r`

5. **crear_entitat**

Entrada: Taula sencera en format taula

Retorna: un conjunt de llistes:

- `capçalera`: Descripció taula
- `models`: nombre de models
- `molècules`: número i nom de les molècules que formen l'estructura quaternària
- llista de residus de la taula mare amb el format definit en el mètode `residut` de la classe Atributs
- llista d'àtoms de la taula mare amb el format definit en el mètode `atomt` de la classe Atributs. A les coordenades dels àtoms se'ls hi sumaran els

resultats obtinguts a partir del mètode cercar_centre _proteïna d'aquesta classe. Per tan amb les coordenades dels àtoms centrades

- En cas que el registre procedeixi de un registre "ATOM " el camp idTAtom prendrà el valor de True si procedeix de un registre "HETATM" prendré el valor de False
- En cas que el registre procedeixi de registres "ATOM " el camp idTResidu prendrà el valor de True si procedeix de registres "HETATM" prendré el valor de False.

Formats de la llista de sortida:

Atoms

<i>idModel,</i>	<i>Número de model</i>
<i>idChain,</i>	<i>Identificador de molècula</i>
<i>idnumSeqres,</i>	<i>Identificador residu</i>
<i>idAtom,</i>	<i>Identificador àtom</i>
sigatom,	Element químic àtom + afegit repetició
nomResidu,	Sigles del residu
xcoor,	X coordenada
ycoor,	Y coordenada
zcoor,	Z coordenada
elemQuim,	element químic
idCA,	True si és un carboni α
idTAtom	True si prové d'àtom, False si és d'un "HETATM"

Residus

<i>idModel</i>	<i>Identificador residu</i>
<i>idChain</i>	<i>Identificador de molècula</i>
<i>idResidu</i>	<i>Identificador residu</i>
nomResidu	Sigles del residu
xcoor	X coordenada
ycoor	Y coordenada
zcoor	Z coordenada
nAtoms	Número d'àtoms
idTResidu	True si prové d'àtom, False si és d'un "HETATM"

6. cercar_centre_proteina

Entrada: registres dels coordenades de la llista d'àtoms

Retorna: 0-coordenades mitjanes de la llista d'àtoms

5.1.2 FUNCIO MAXMINS

Entrada: llista d'àtoms

Retorna: valors màxims, mínims i centrals de tots els àtoms

5.2 CODI.ATRIBUTS

El mòdul atributs conté dues classes, Colors i Atributs.

5.2.1 CLASSE COLORS

La classe colors pot rebre dos paràmetres(ambdós no obligatoris)

5.2.1.1 ATRIBUTS

1. **taula_colors** llista amb la cadena de colors que formen la paleta en format (r,g,b, α)
2. **molècules**. Diccionari de totes les lletres en majúscula que puguin utilitzar-se per anomenar molècules com a clau i el número del color de la taula_colors
3. **molecules2**. Diccionari similar a l'anterior però amb el color codificat per utilitzar-se en el matplotlib
4. **tipus_aminoacid**, Diccionari de sigles dels aminoàcids com a clau i un número de l'1 al 5 segons l'atribut següent
5. **tipologia_amin** diccionari en que la clau son els números del 1 al 5 i el resultat de l'1 al quatre és si l'aminoàcid es hidrofòbic, hidrofílic, àcid o bàsic el número 5 correspon a residus que no són d'aminoàcids
6. **tipologia_lcolor**. Diccionari igual que l'anterior però que el resultat es un lletra minúscula per utilitzar en el matplotlib
7. **tipologia_color** Diccionari igual que l'anterior però que el resultat es un número de color de la taula_colors
8. **color_aminoacids**. Diccionari clau: Aminoàcid (sigles 3 lletres) el valor és el número de color de l'atribut **taula_colors**

5.2.1.2 MÈTODES

1. **color_molecula ()**

Entrada: parametre1 i paràmetre2 de la classe (valors 1 o 2) el paràmetre 1 és la lletra de la molècula a acolorir i si el paràmetre2 és 1

Retorna el color de l'atribut **taula_colors** i si es 2 retorna el resultat de l'atribut **molècules**

2. **color_tipus()**

Entrada: sigles aminoàcid en el parametre1 i parametre2 (1=3D , 2=2D)

Retorna: atribut de color. Si parametre2==1 **tipologia_lcolor** color(r,g,b,α)

Si parametre2==2 **tipologia_color** color(lletra)

3. **color_aminoacid()**

Entrada: parametre1 = sigles aminoàcid (3 lletres)

Retorna: color(r,g,b,α) corresponent al atribut **taula_colors**

5.2.2 ATRIBUTS

La classe Atributs com la classe Colors pot rebre dos paràmetres(ambdós no obligatoris)

5.2.2.1 ATRIBUTS

1. **atom_radis** diccionari que la clau són el elements químics presents en el cos humà i el valor és el radi

2. **mestre_aminoàcids** llista d'atributs dels 20 aminoacids amb els camps. Sigles (3 lletres), nom sencer, sigla (1 lletra), massa del aminoàcid, carrega , tipus d'aminoàcid, relació de la massa respecte a la Glicina que té el valor 1 màxim triptòfan 3,4.

5.2.2.2 MÈTODES

1. **campAtoms()**

Entrada: parametre1 == nom del camp de la llista àtoms

Retorna: l'índex de la llista a que correspon el camp

2. **atomt()**

Entrada: parametre1 == nom del camp de la llista atoms

Retorna: l'índex de la llista a que correspon el camp

3. residut()

Entrada: parametre1 == nom del camp de la llista residus

Retorna: l'índex de la llista a que correspon el camp

4. radi()

Entrada: parametre1 == sigles d l'element químic de l'àtom

Retorna: valor del radi en armstrongs

5. nom_ aminoacid()

Entrada: parametre1 == sigles de l'aminoàcid (3 lletres)

Retorna: el nom de l'aminoàcid tret de la llista mestre_ aminoacids

6. llista_ aminoacids()

Entrada: cap

Retorna: la llista mestre_ aminoacids en forma de taula

5.3 CODI.DADES

5.3.1 CLASSE DADES

Rep el paràmetre de la taula a resumir

5.3.1.1 ATRIBUT

Es pot considerar com a dades d'entrada el resultat del mètode **crear_entitat** de la classe Entitat que te com a resultat, com s'ha vist anteriorment, les dades descriptives, les molècules, els residus, els àtoms i el propi nom de la taula.

5.3.1.2 MÈTODES

1. entrades()

Entrada (self)

Retorna: capçalera, molècules, nom_taula

2. get_header()

Entrada (self)

Retorna: text amb el conjunt de línies "HEADER"

3. get_helix

Entrada (self)

Retorna: text amb el nombre de registres "HELIX "

4. **get_sheet()**

Entrada (self)

Retorna: text amb el nombre de registres "SHEET " i el nombre de "SHEETS " amb codi diferent

5. **get_title()**

Entrada (self)

Retorna: text amb el conjunt de línies de format "TITLE "

6. **get_models_dades()**

Entrada (self)

Retorna: text amb el nombre de registres "MODEL " en cas que no hi gai retorna 1

7. **get_keywds()**

Entrada (self)

Retorna: text amb el nombre de registres "KEYWDS

8. **get_compnd()**

Entrada (self)

Retorna: llista amb registres de format "COMPND" definició de les molècules

9. **get_heterogens()**

Entrada (self)

Retorna: llista amb el conjunt de elements heterògens de que inclou els "HET ", el nombre d'àtoms de "HET ", el registre "HETNAM", els registre "HETSYM" i els registres "FORMUL"

10. **get_natoms()**

Entrada (self)

Retorna: el número d'àtoms del complex

11. **get_nresidus()**

Entrada (self)

Retorna: el número de residus del complex

12. **get_elements()**

Entrada (self)

Retorna el numero d'àtoms segons elements de la proteïna i el nombre d'àtoms segons element dels registres heterògens

13. `get_residus()`

Entrada (self)

Retorna el numero de residus segons elements de la proteïna i el nombre de molècules d'heterògens segons element dels registres heterògens i el nombre de cadascun del residus de la proteïna en cas que no es tracta d'aminoàcids

5.4 PLOT2D.PROTEINA2D

L'ús de la llibreria Matplotlib per representar l'estructura de les proteïnes no dóna un resultat massa espectacular comparat amb el que podem veure amb el panda3D, això no obstant ens dóna una idea de l'estructura.

La dificultat per utilitzar formes geomètriques tret dels marcadors típics, és possible però la llibreria que vaig trobar només funcionava per Windows. Per tant s'han escollit tres tipus de gràfics 3D.

5.4.1 CLASSE PLOTS

Paràmetres que rep:

tipus (obligatori) valors a (punts: àtoms residus) o c (línies)

tipus2 (obligatori) en cada mètode s'explicarà

5.4.1.1 ATRIBUTS

Es pot considerar com a dades d'entrada el resultat del mètode `crear_entitat` de la classe `Entitat` que te com a resultat, com s'ha vist anteriorment, les dades descriptives, les molècules, els residus, els àtoms i el propi nom de la taula.

5.4.1.2 MÈTODES

1. `plot_atoms(self)`

Es tracta d'un scatterplot 3D

En aquest cas si `tipus == 1` sempre és `a` perquè sempre es representen punts

Tipus2 :

- molècula àtoms acolorits per molècules
- residus representat com àtoms i acolorits per molècules

en aquest cas només es representa l'acolorit segons molècules, però afegint un tercer paràmetre es podria canviar per qualsevol tipus d'atribut, tipologia d'aminoàcid, àtoms per color de l'aminoàcid etc.

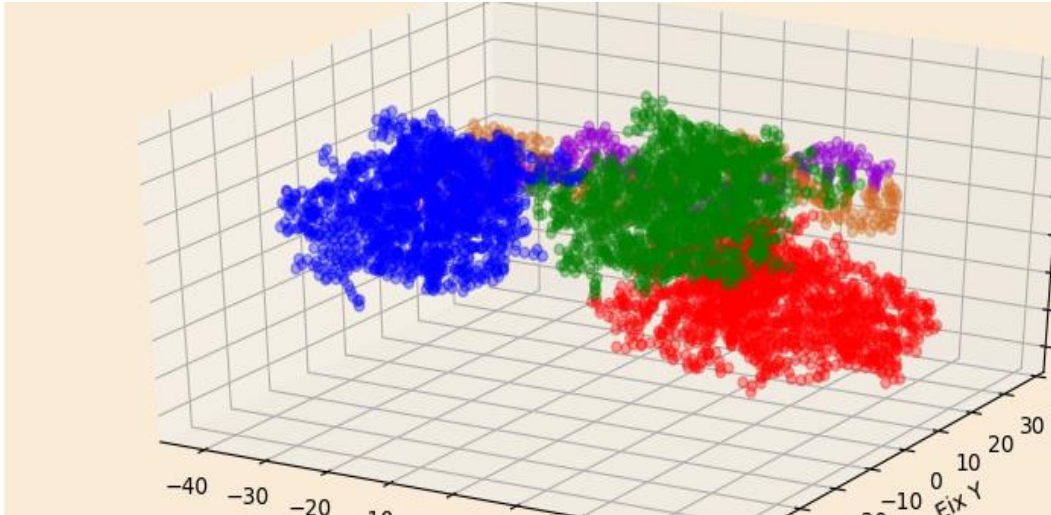


Figura 24. Proteïna 1TUP per àtoms i molècules

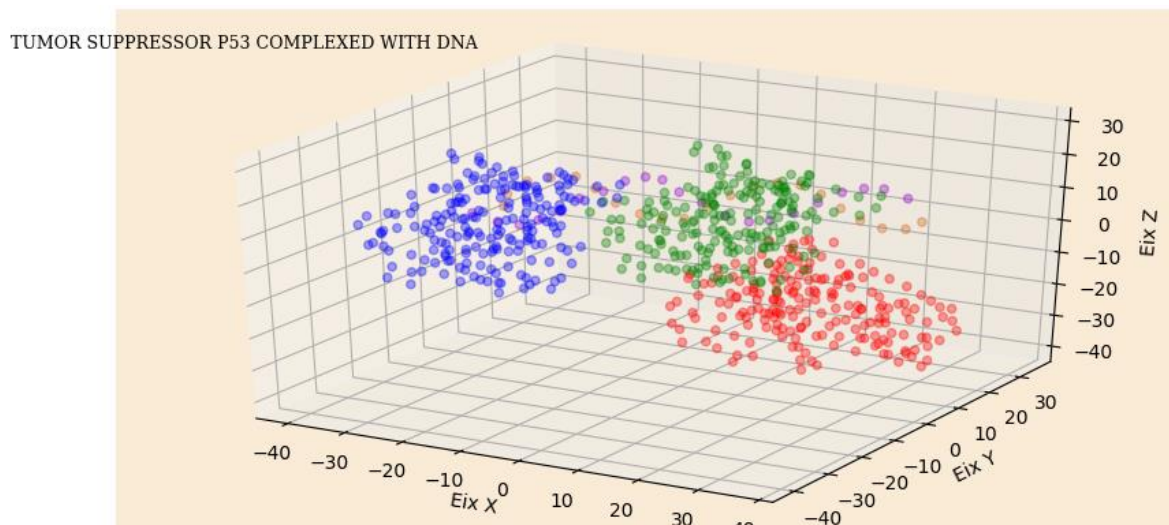


Figura 25. Proteïna 1TUP per residus i molècules

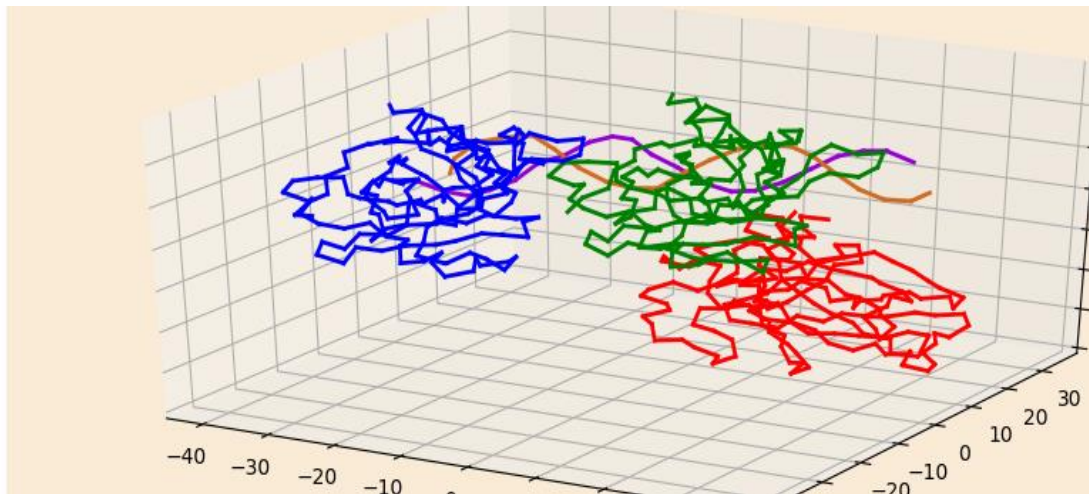
2. plot_linies (self)

En aquest cas es tracta d'unir els carbonis α amb línies que representarien l'aminoàcid o el tros de cadena d'ADN.

En aquest cas si tipus ==1 sempre és c perquè sempre es representen línies

Tipus2 :

- línies acolorides per molècules
- línies acolorides per tipologia aminoàcid



• Figura 26. Proteïna 1TUP per línies(aminoàcids) i molècules

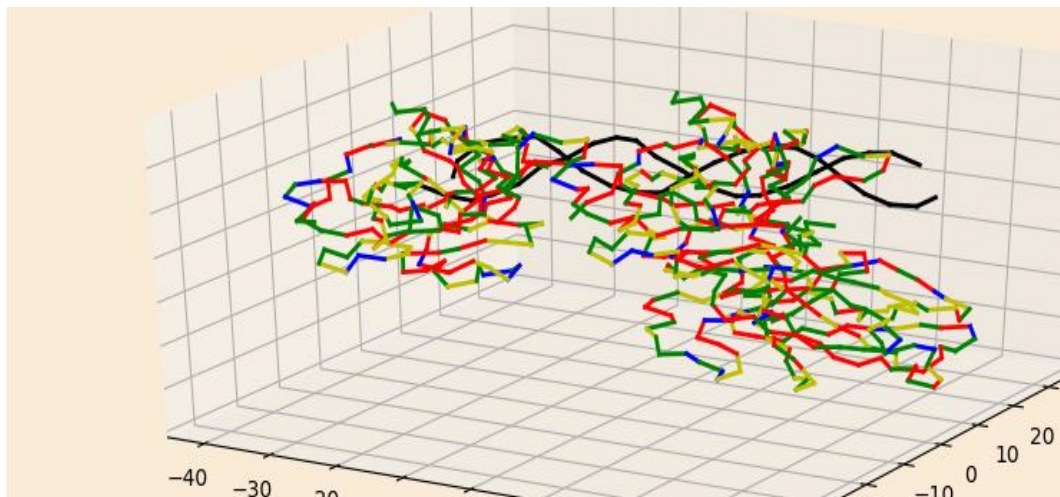


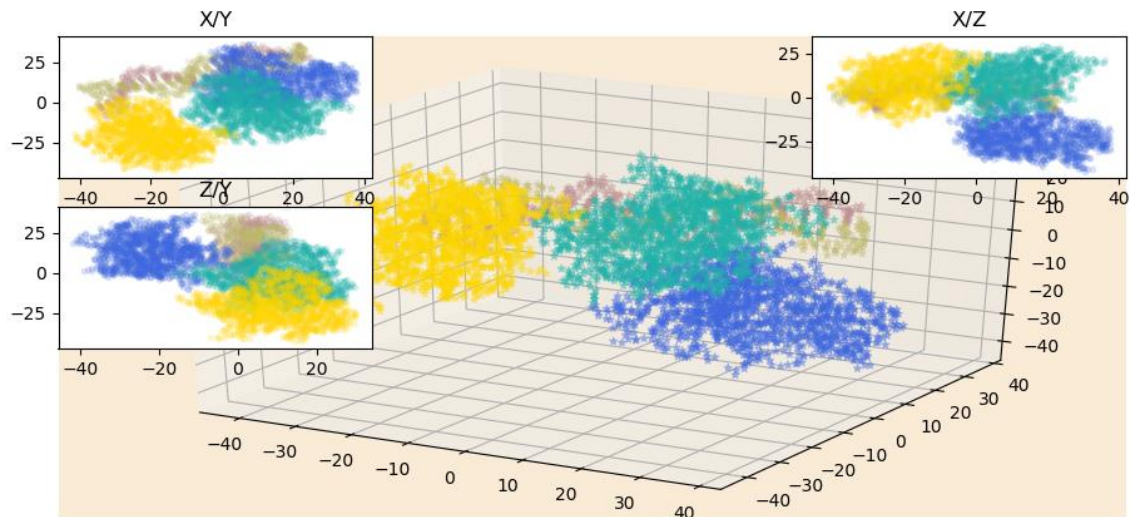
Figura 27. Proteïna 1TUP per línies i tipus aminoacids

3. plot2d-3d

En aquest cas es tracta de tractar de representar 4 vistes amb un sol gràfic, Una amb 3d i la projecció d'aquesta sobre els tres plans

El tipus1 = sempre és **a** perquè sempre es representen 'a'

El tipus2 : sempre es 'vistes'



TUMOR SUPPRESSOR P53 COMPLEXED WITH DNA

Figura 28. Proteïna 1TUP diferents vistes

5.5 CODI.PLOT3D.PROTEINA3D

5.5.1 CLASSE PROTEINES

En el Panda3D existeix una variable temps que és una instància de la classe ClockObject. Per obtenir el temps (en segons) des de l'última trama: s'utilitza la fórmula $dt = \text{globalClock.getDt}()$. Aquesta és la base del moviment utilitzant dt multiplicat pels atributs definits per l'usuari:

5.5.1.1 ATRIBUTS

El multiplicador pels desplaçaments:

Lineal: $\text{pas_mov} = 15$

Gir: $\text{pas_rot} = 20$

Ample de la finestra: 1000 punts

Alçada de la finestra: 800 punts

Color de fons

Paràmetres de la classe:

- Tipus: Implica quin tipus de representació:
 - Tipus = 1, 2, 4, 5, 7 representació per atoms
 - Tipus= 3 representació tipus ball & stick
 - Tipus = 7 representació tipus backbone

- Fitxer
- Amino (optatiu) l'aminoàcid a representar dins la proteïna

Com en altres classes s'utilitza la classe Entitat i el mètode crear entitat que retorna la informació en forma de llistes:

- Capçalera
- Molècules
- Residus
- Àtoms
- Nom_taula

Obtenir de codi.models

- Atom_sphere.egg Imatge del àtom
- Dirlight.egg Llum direccional

Es crea el node de panda3D proteïna :

```
self.proteina = render.attachNewNode("Proteina")
```

5.5.2 METODES

5.5.2.1 DIBUIX

1. **atomic**

Rep el node proteïna com a paràmetre amb el nom genèric de node va dibuixant el model Atom_sphere.egg, mitjançant el fitxer àtoms s'obté el radi amb el mètode Atributs.radi() vist anteriorment passen com a paràmetre el codi de l'element i després s'acolorix el conjunt d'àtoms que formen un aminoàcid amb el color procedent de la funció inclosa en la aquesta classe **ccolor()** que veurem posteriorment

Diferència el tipus 4 dels demes a efectes d'acolorir els àtoms. En el cas del tipus =4 utilitza el mètode Atributs.nom_aminoacid() per poder presentar el nom del aminoàcid entrat.

Finalment es dibuixen els tres eixos cartesianes que passen pel centre de la proteïna i que s'afegeixen al **node**

2. ball_stick

En aquest cas també rep el node que s'ha creat al principi. Aquest mètode farà servir un sistema de dibuix mixt ja que es dibuixen els carbonis α i alhora les línies que simbòlicament els uneixen. no són res més que l'espai que ocupant la resta d'element de l'aminoàcid. En aquest cas es genera jeràrquicament segons el model de dades

Per cada molècula de la llista molècules es selecciona mitjançant una *list comprehension*

```
Linies = [linies for linies in atoms
           If linies[self.c] == idCadena and
           linies[self.prot] == True and
           linies[self.sig] == "CA"]
```

idCadena = identificador de molècula
prot = True que no sigui heterògen
i que sigui un carboni α

per tan la llista línies només hi hauran els carbonis alfa de la proteïna i de la molècula N_1 de les N molècules que formen la proteïna .

En aquest cas el carboni α s'acolorirà amb el color que li correspon a la tipologia de l'aminoàcid que pertany mitjançant el mètode Color.color_tipus amb format LVecBase4f , per fer-ho més estàndard es busca el radi de l'element (en aquest cas tots són carbonis i seria innecessari) mitjançant la classe Atributs.radi() i es dibuixa el model atom_sphere.egg

Es crea un node lins = Linesegs(). Per cada registre de la llista es graven les coordenades a una nova llista anomenada coordenades. Un cop dibuixats tots els àtoms de la molècula mitjançant un bucle és fa

En el primer registre es fa un move.To (x_0, y_0, z_0) i a partir d'aquest consisteix en fer un draw.to(x_n, y_n, z_n) un cop acabat se li donen els atributs al gruix de línia, color i es crea el node de línies lnode = lins.create() i s'inclou en el node general.

Es dibuixen els eixos de coordenades per la referència.

3. **backbone**

La funció backbone,, realment no representa un backbone, ja que el dibuix d'un sistema tipus "tubular" no és pot fer(que jo sàpiga) en panda3D, ja que seria necessària uns elements cilindre.egg que tingues la possibilitat o be de coordenades inicials i coordenades finals i el radi i el color, o bé el punt inicial i els angles α i β del pla x respecte als plans y i z, igualment un element colze.egg que servis per ajuntar els dos tubs(cilindres)

El fet que no existeixin aquests possibilitats (tot i que no em dono per vençut i intentaré mitjançant blender aconseguir aquests dos models que no existeixen .

Per tant, en absència, dels models necessaris s'han utilitzant les línies de major gruix.

Per tant el sistema es agafar la llista de coordenades dels carbonis α i començar el primer registre per move.To (x_0, y_0, z_0) i el següents per draw.To (x_a, y_a, z_a) fins el darrer registre de la molècula i començar el registres de la propera molècula de la mateixa manera que en la primera molècula fins l'última.

5.5.2.2 MÈTODES DE MOVIMENT

4. **tecles_accio**

Definició de les tecles que s'utilitzaran per activar els moviments de rotació o d'apropar o allunyar

5. **moviments**

En aquesta funció s'estableixen les accions a realitzar quan es prem una de les tecles escollides es tracta de augmentar o disminuir la quantitat dt

$$\text{pas} = \text{Proteines.pas_mov} * dt$$

$$\text{rot} = \text{Proteines.pas_rot} * dt$$

Z: apropar

X: allunyar

E: rotar en diagonal avall

R: rotar en diagonal amunt

fletxa amunt: girar sobre l'eix x girar cap avall

fletxa avall: girar sobre l'eix x girar cap amunt

fletxa avall: girar sobre l'eix z girar cap esquerra

fletxa avall: girar sobre l'eix z girar cap a la dreta

6. **centrar_camera**

Situa en el centre l'eix xz i se situa del centre de coordenades en el eix y a una distància igual a dues vegades al radi de l'envoltant de la proteïna

7. **centrar_proteïna**

Tot i que centraren la classe entitat es cerca el centre de totes les coordenades i es suma o resta la quantitat a tots els àtom perquè el centre de la proteïna se centra en el punt 0,0,0. Quan es genera la proteïna es torna cercar el centre de coordenades hi ha una petita diferència que se soluciona en aquesta funció.

8. **maxims_minims**

Entrada: Els àtoms que formen part de la proteïna

Es cerquen els valors màxims i mínims de cada eix per poder saber els màxims i mínims per dibuixar l'eix de coordenades en $z_{min} - n$ i $z_{max} + n$

Entrega: x_{max} , y_{max} , z_{max} , x_{min} , y_{min} , z_{min}

5.5.2.3 ALTRE CODI INCLÒS EN EL COS DEL PROGRAMA

1. Il·luminació

S'utilitzen diferents tipus d'il·luminació amb la finalitat de donar l'efecte de les dimensions.

- Llum ambiental que il·lumina l'escena en general
- Llum direccional 1 es situa en una determinada posició el llum amb una orientació determinada i un color determinat
- Llum direccional 2 es situa en una posició apartada de l'anterior formant un determinat angle sobre la seva ubicació de la càmera

- Llum spotlight es tracta d'una il·luminació puntual amb unes determinades característiques especialment la posició (x, y, z) i el punt a que apunta en coordenades en aquest cas el centre de coordenades

5.5.3 EXEMPLES GRÀFICS

OPCIO

B01.Representacio del complex proteic segons color aminoàcids

B02.Representacio dels elements externs a la proteïna

B03.Representacio per tipus aminoàcids

B04.Representacio ubicació d'un aminoàcid concret

B05.Representacio del complex per molècules

B06.Representacio ball & stick els acolorit per tipus aminoàcid

B07.Representacio tipus backbone

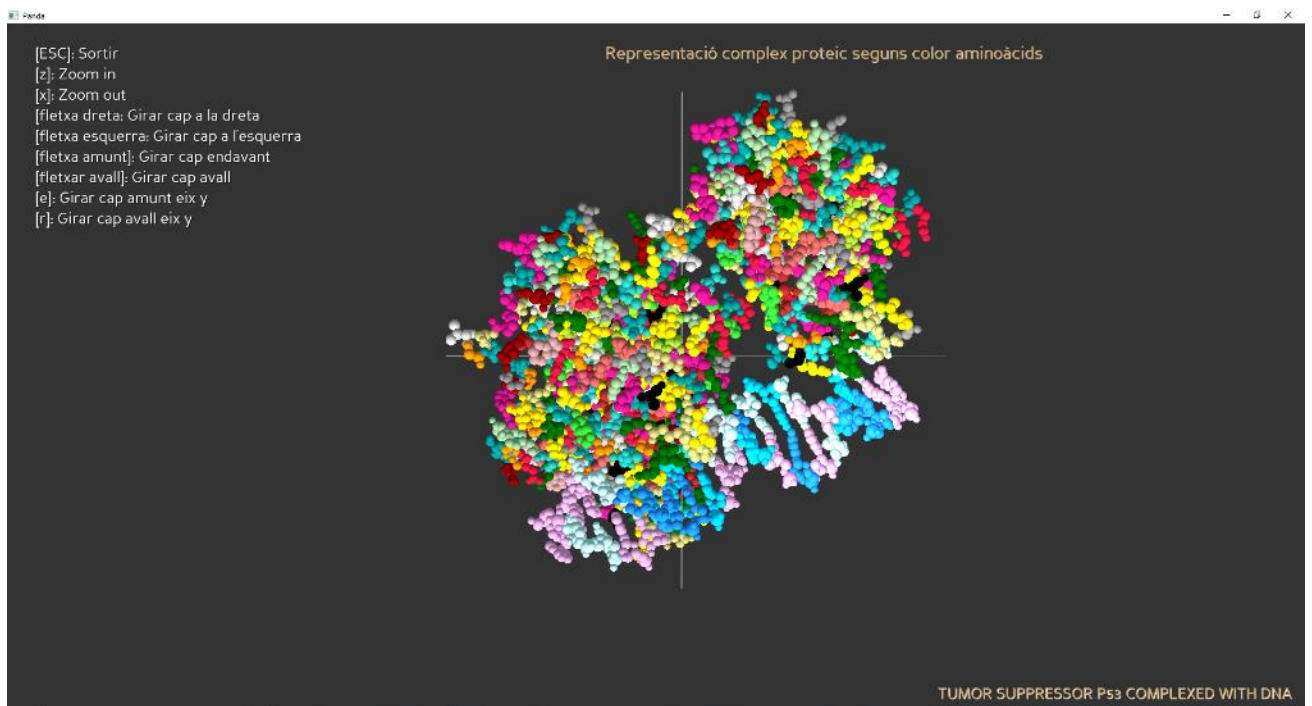


Figura 29.- .Representació del complex proteic segons color aminoàcids

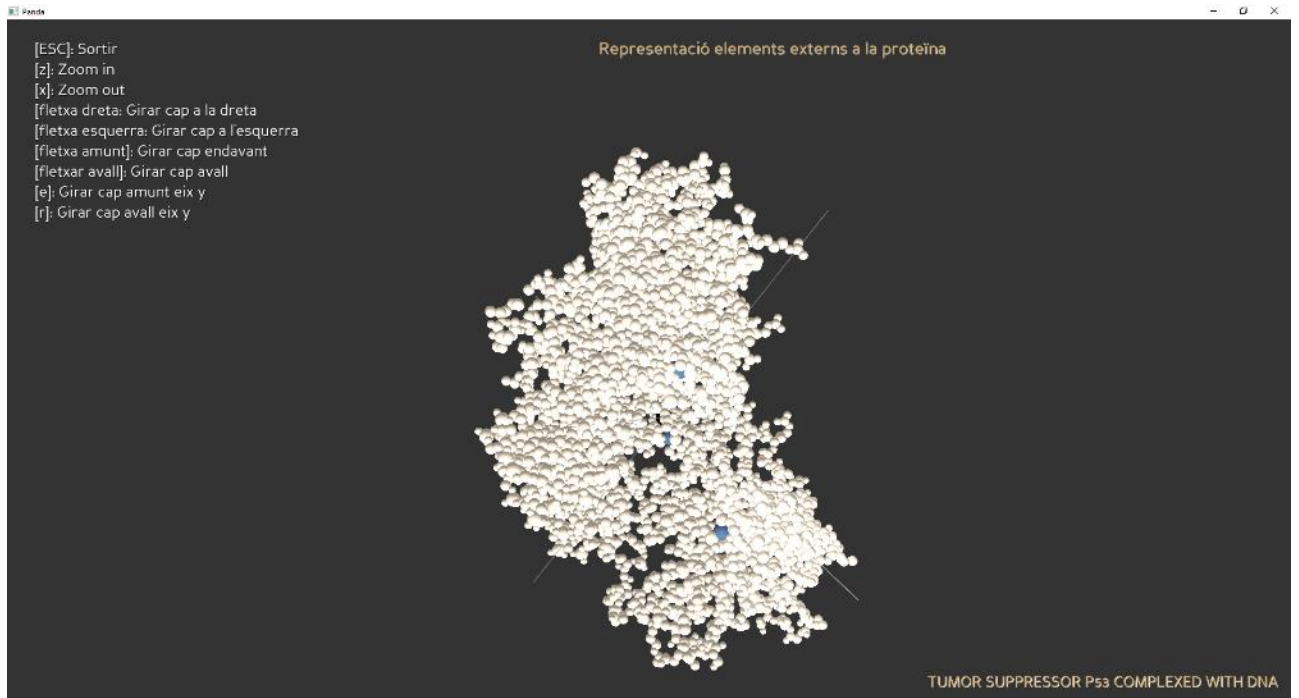


Figura 30.- Representació dels elements externs a la proteïna

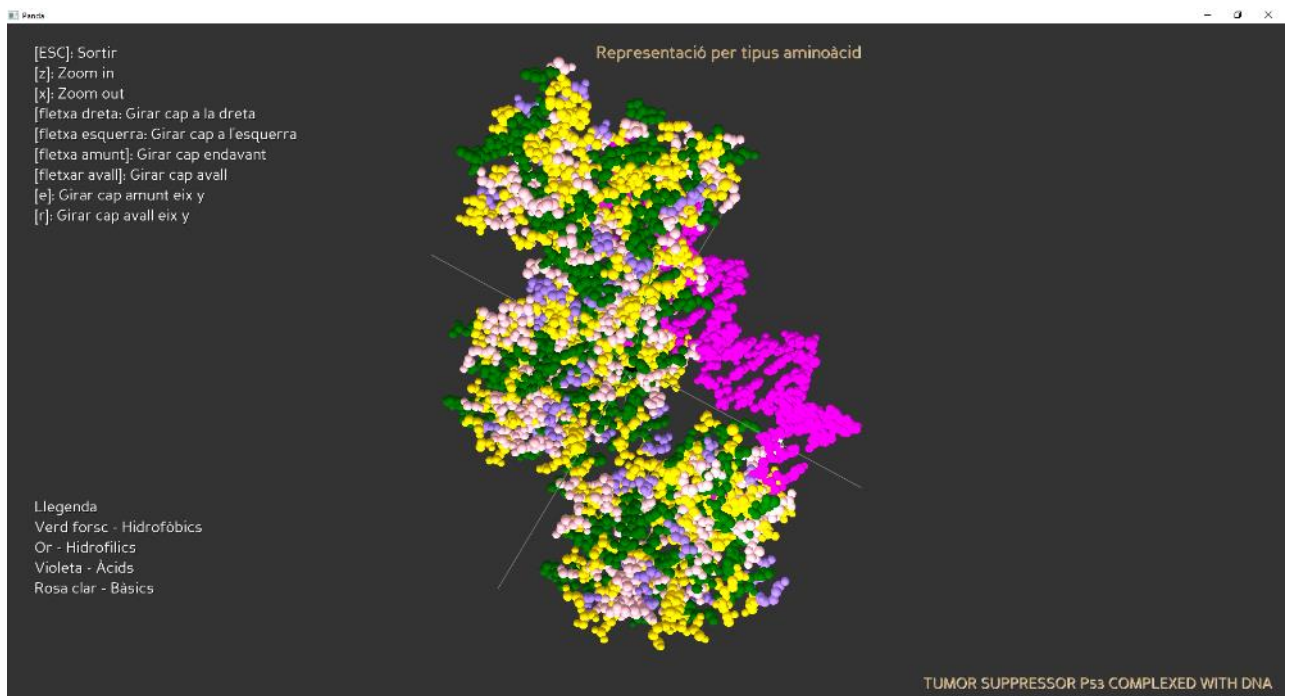


Figura 31.- Representació dels elements externs a la proteïna

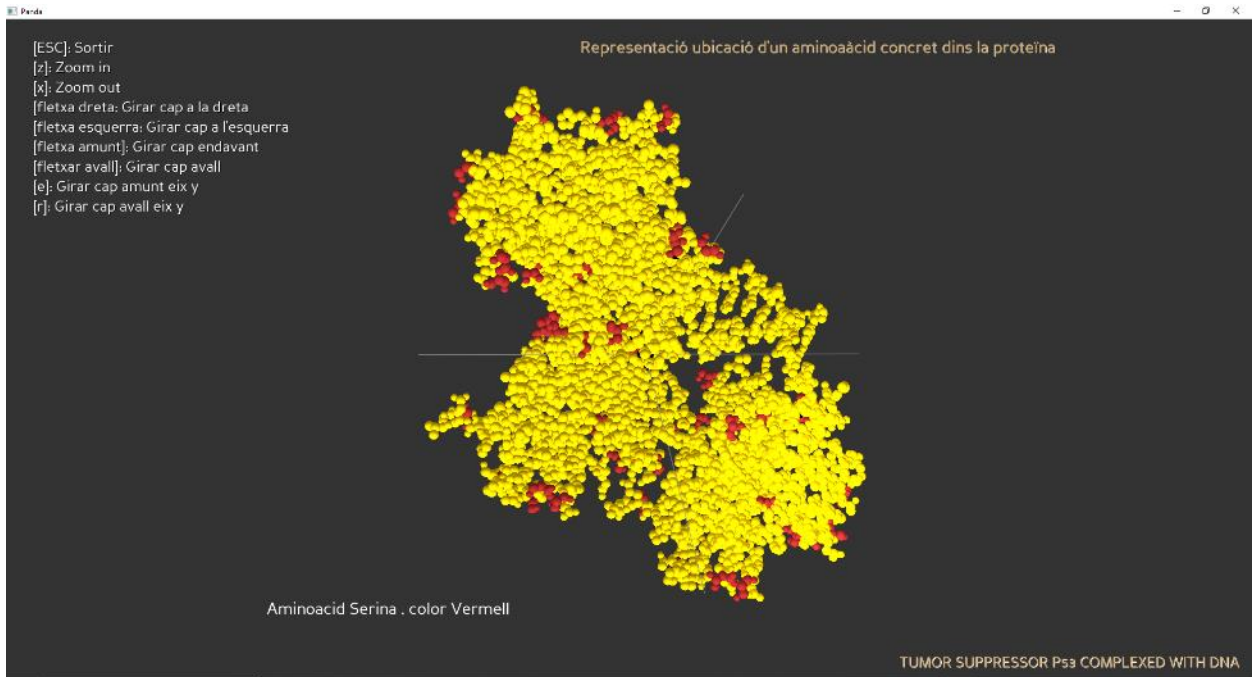


Figura 32 Representació ubicació d'un aminoàcid concret

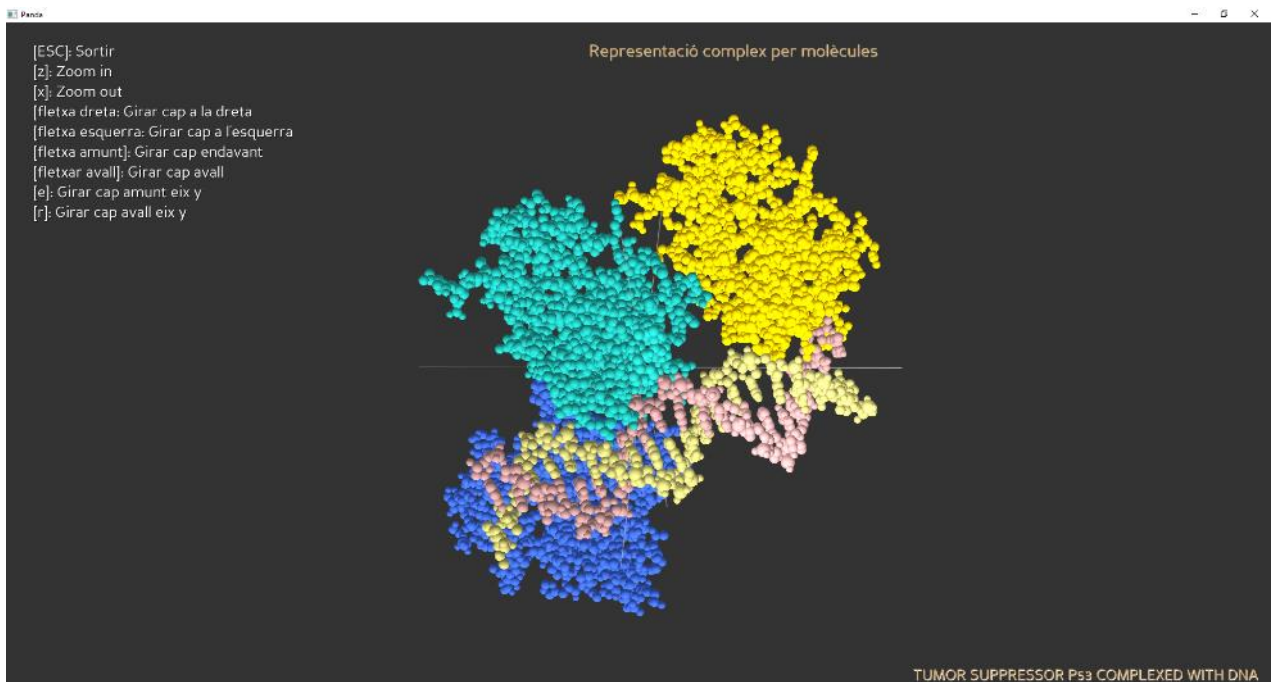


Figura 33 .- Representació del complex per molècules

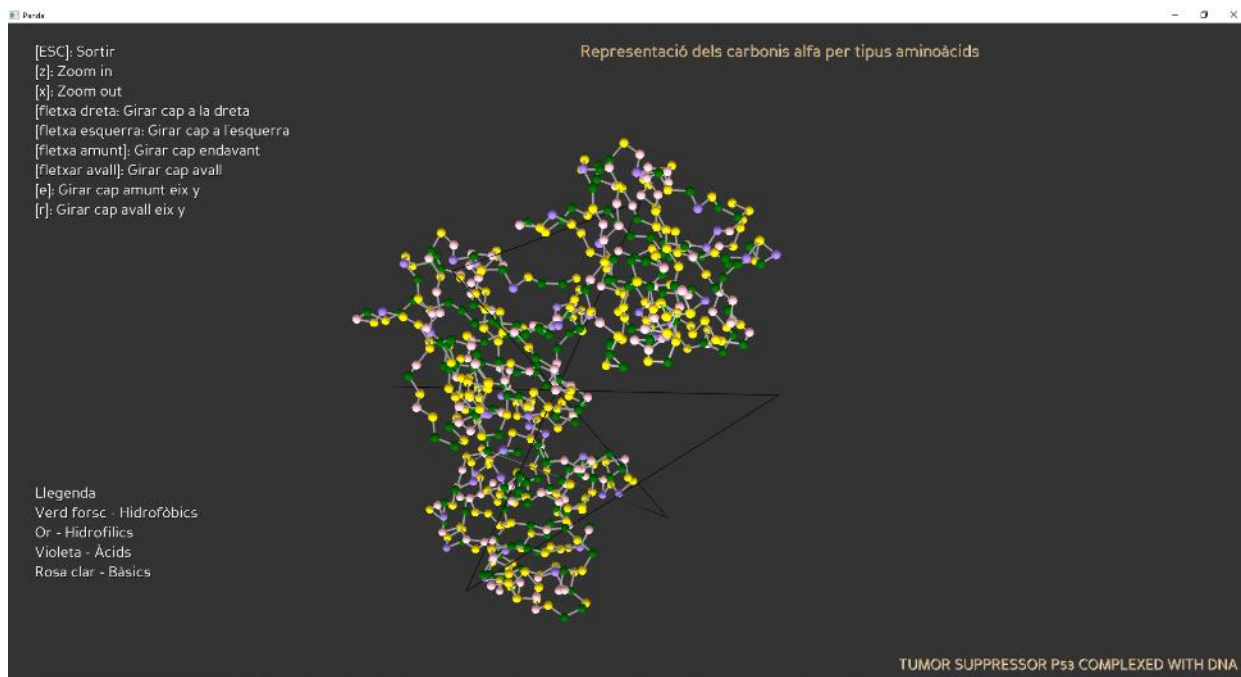


Figura 34.- Representació ball & stick els acolorit per tipus aminoàcid

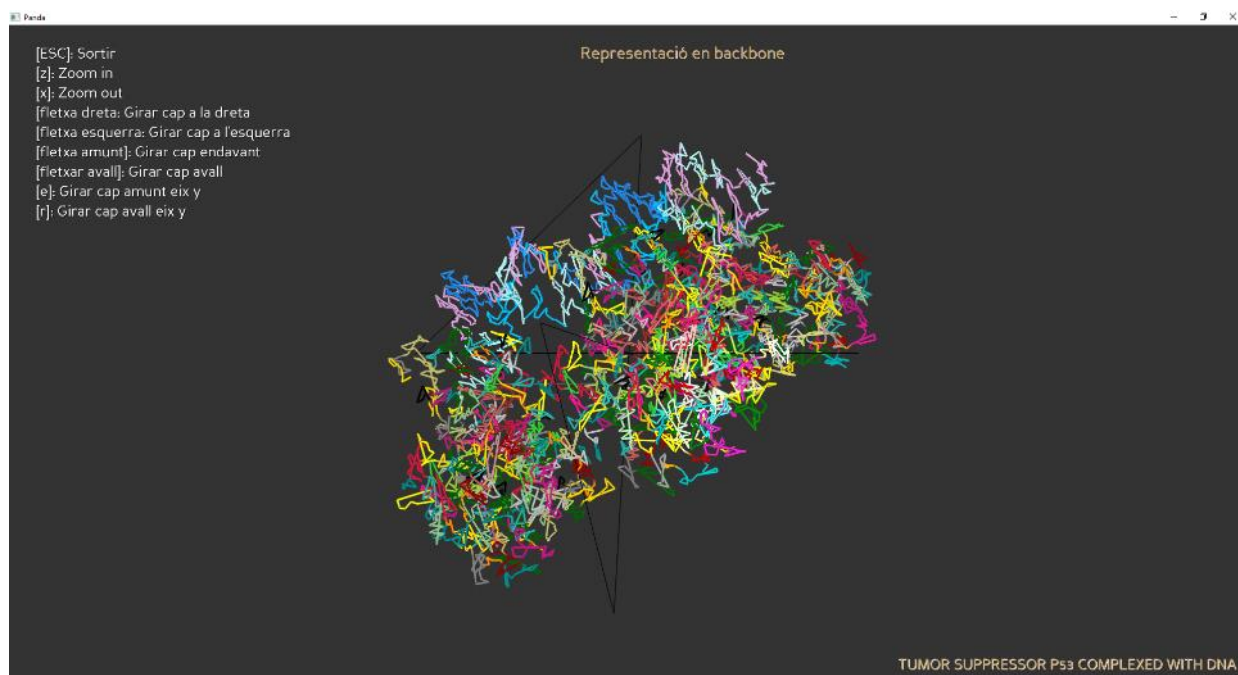


Figura 35.- Representació tipus backbone

5.5.4 SOBRE LA RENDERITZACIÓ

La renderització de les representacions no atòmiques no han donat problemes en cap cas, això no obstant, existeixen limitacions en la renderització quan es tracta de la representació amb àtoms, que depèn del maquinari utilitzat i del nombre de residus a renderitzar, així s'ha provat el mòdul de 3D per la representació en àtoms en tres plataformes diferents. Dissortadament no he pogut comprovar-ho exactament amb el mateix maquinari. Les plataformes testades són:

Sistema operatiu	Marca	Processador	Ghz	GB Ram
Lubuntu 16.04 LTS	DELL	i7	?	8
OSX Gran Capitan	MAC BOOK PRO	i5	2,36	8
Windows 10	HP	i7	2,66	16

S'han provat 4 fitxers dels exemples amb àtoms i s'han trobat limitacions i diferents temps de renderització:

Proteïna	Residus	Lubuntu	Mac OSX	PC
3J2T	8750	1'26,00"	2'21,95"	2'34,60"
3G37	4632	45,37"	1'05,28"	1'52,61"
3J6R	2868	23,29"	29,28"	46,33"
1TUP	630	8,50"	9,20"	15,00"

Es pot veure clarament que el SO millor és la de Linux Ubuntu i concretament la distribució Lubuntu 16.04 Xenial LTS en totes les proteïnes provades amb uns temps de renderització quasi-lineals amb una regressió polinòmica el paràmetre de x^2 és de $3 \cdot 10^{-7}$ essent x el nombre de residus. El Mac funciona en totes les proteïnes però l'increment es menys lineal que l'anterior ja que el paràmetre de x^2 és de $7 \cdot 10^{-7}$, en el cas del PC de les 4 proteïnes la de 2868 residus a part de tardar el doble que les plataformes anteriors ja no es renderitzen els textos, a la de 4.632 residus només renderitza una part mínima d'àtoms i es queda com si hagués acabat, amb la de 8750 residus directament al cap de més de 2 minuts el programa "peta".

Per analitzar més detalladament el cas de Windows s'ha provat la proteïna de 1206 residus i funciona correctament, és a dir, entre 1.200 i 2.868 residus en Windows 10 hi ha un punt en que ja no renderitza correctament

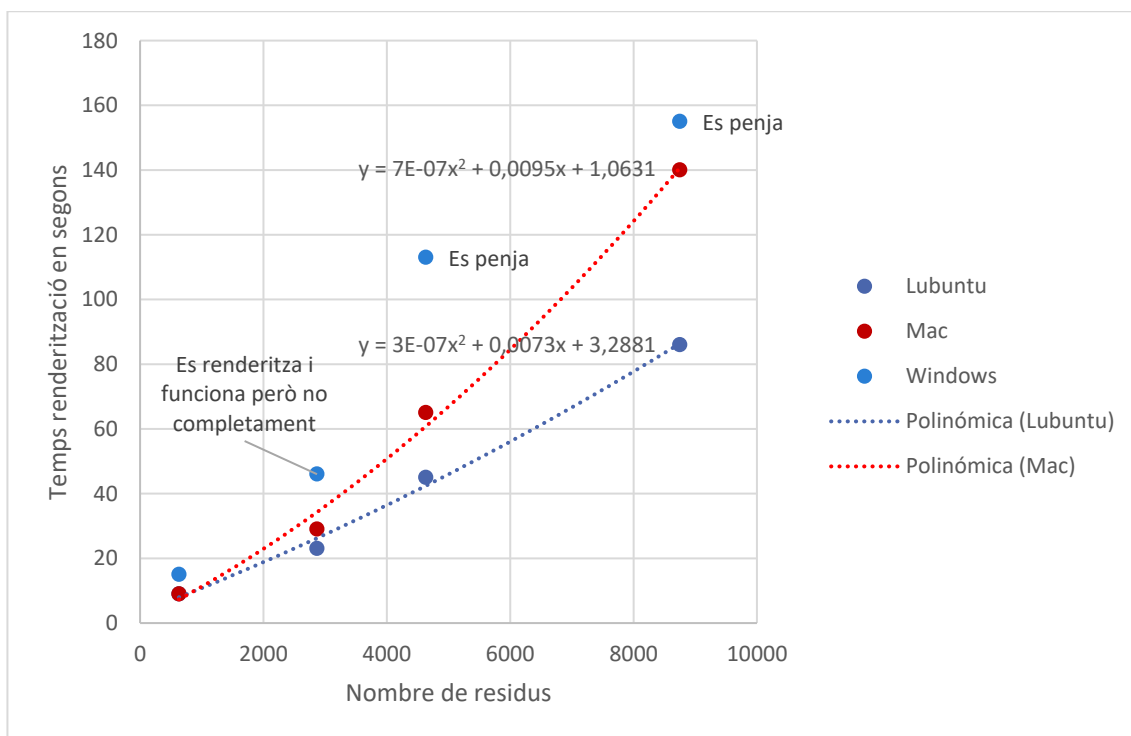


Figura 36. Renderització segons plataforma i nombre de residus en el cas de representació atòmica

6. RESTA DE MODULS

La resta de mòduls que componen el projecte són bàsicament els que efectuen les crides a les diferents classes o presenten plans

6.1 .GITIGNORE

Es un mòdul auxiliar que serveix per filtrar els fitxers que no s'han de pujar al GitHub

6.2 AMINOACIDS.PY

Es un mòdul d'ajuda que recorda la informació dels aminoàcids, especialment per si és vol fer una representació en el visor 3D de la opció B04.Representacio ubicació d'un aminoàcid concret el resultat és el següent:

Aminoàcid	SIG	Lletra	Tipologia	Relació massa (*)
Alanina	A	ALA	1	1,40
Arginina	R	ARG	4	3,08
Asparagina	N	ASN	2	2,00
Aspartato	D	ASP	3	1,90
Cisteína	C	CYS	2	1,79
Glutamato	E	GLU	4	2,27
Glutamina	Q	GLN	1	2,38
Glicina	G	GLY	2	1,00
Histidina	H	HIS	2	2,46
Isoleucina	I	ILE	4	2,58
Leucina	L	LEU	1	2,58
Lisina	K	LYS	1	2,81
Metionina	M	MET	2	2,58
Fenilalanina	F	PHE	3	2,81
Prolina	P	PRO	1	1,88
Serina	S	SER	2	1,52
Treonina	T	THR	2	1,94
Triptófano	W	TRP	1	3,40
Tirosina	Y	TYR	2	2,94
Valina	V	VAL	1	2,19

* massa relativa a la massa de la Glicina = 1

Tipologia	Descripció
1	Hidrofòbics
2	Hidrofílics
3	Àcids
4	Bàsics

6.3 COLORS.PY

Ens mostra la paleta de colors usada:



6.4 FITXERS

Es un mòdul que serveix per escollir el fitxer amb que treballar, per evitar entrar-lo com a paràmetre cada vegada que es vulgui obtenir un tipus de gràfic. El fitxer escollit es guarda en la taula taula.idx i no es canvia fins que es torna a utilitzar el mòdul

Escollir un fitxer			Num. Residus
1	1acc.pdb	Anthrax protective antigen .	667
2	1bkv.pdb	Collagen .	94
3	1e9y.pdb	Crystal structure of helicobacter pylori urease in complex withacetohydroxamic acid .	810
4	1tup.pdb	Tumor suppressor p53 complexed with dna .	630
5	1yd9.pdb	1.6a crystal structure of the non histone-domain of thehistone variant macroh2a1.1. .	754
6	3g37.pdb	Cryo-em structure of actin filament in the presence of phosphate .	4632
7	3j6r.pdb	Electron cryo microscopy of human papillomavirus type 16 capsid .	2868
8	3pte.pdb	The refined crystallographic structure of a dd-peptidase penicillin target enzyme at 1.6 a resolution .	347
9	3tnu.pdb	Heterocomplex of coil 2b domains of human intermediate filamentproteins, keratin 5 (krt5) and keratin 14 (krt14) .	185
10	3wct.pdb	The structure of a deoxygenated 400 kda hemoglobin provides a moreaccurate description of the cooperative mechanism of gianthemoglobins: oxygenated form	1204
11	4bj3.pdb	Integrin alpha2 i domain e318w-collagen complex	421
12	4dkl.pdb	Crystal structure of the mu-opioid receptor bound to a morphinanantagonist	460
13	4nco.pdb	Crystal structure of the bg505 sosip gp140 hiv	2841
14	5ha9.pdb	Crystal structure 1 env trimer in complexwith the broadly neutralizing fab pgt122 pa10) that induces apoptosis and inhibits metastasisin triple negative breast cancer	711
15	5I6X.pdb	Ray structure of the ts3 human serotonin transporter complexed withparoxetine at the central site .	983
16	5uru.pdb	Insulin with proline analog dhp at position b28 in the r6 state .	206
17	6env.pdb	X-ray structure of au2phen-encapsulated horse spleen apoferritin .	187

Opció :

6.5 DADES_BAT.PY

Ofereix una fitxa bastant completa de la proteïna estudiada. Aquesta fitxa ja s'ha vist anteriorment, per tant, no fa falta tornar-la a presentar

6.6 OPCIONS.PY

Mòdul d'ajuda que serveix per recordar els paràmetres a introduir a la comanda online per obtenir la presentació desitjada.

	OPCIO 2D i pseudo 3D	Matplotlib	Param 1	Param 2
A01	Representació en pseudo 3d dels àtoms	acolorits per molècules	a	molècula
A02	Representació en pseudo 3d dels residus (carbonis alfa)	per molècules	a	residu
A03	Representació en 2D i pseudo 3D		a	vistes
A04	Representació en línies per molècules en pseudo 3D		c	molècula
A05	Representació en línies dels residus en pseudo 3D segons tipus aminoàcid		c	tipus

	OPCIO 2D i pseudo 3D	Matplotlib	Param 1	Param 2	Param 3
B01	Representació del complex proteic segons color aminoàcids		a	residu	
B02	Representació dels elements externs a la proteïna		a	extern	
B03	Representació per tipus aminoàcids		a	tipus	
B04	Representació ubicació d'un aminoàcid concret		a	amino	sigles_amino
B05	Representació del complex per molècules		a	molècula	
B06	Representació ball & stick els acolorit per tipus aminoàcid		c	alfa	
B07	Representació tipus backbone		c	tipus	

6.7 VISOR2D

El mòdul crida a les representacions 2D

```
>>> python visor2D.py Param1 Param 2
```

6.8 VISOR3D

El mòdul crida a les representacions 3D

```
>>> python visor3D.py Param1 Param 2 [Param 3]
```

El paràmetre 3 es optatiu i només s'utilitza en la opció B04

7. CONCLUSIONS

Una de les primeres conclusions del treball és que la representació gràfica de les proteïnes a partir dels fitxers pdb, a partir del programa clàssic de gràfics utilitzat en python que és el Matplotlib ens dóna unes representacions bastant pobres, tot i que ofereix una idea de la forma de la proteïna per exemple el colàgen

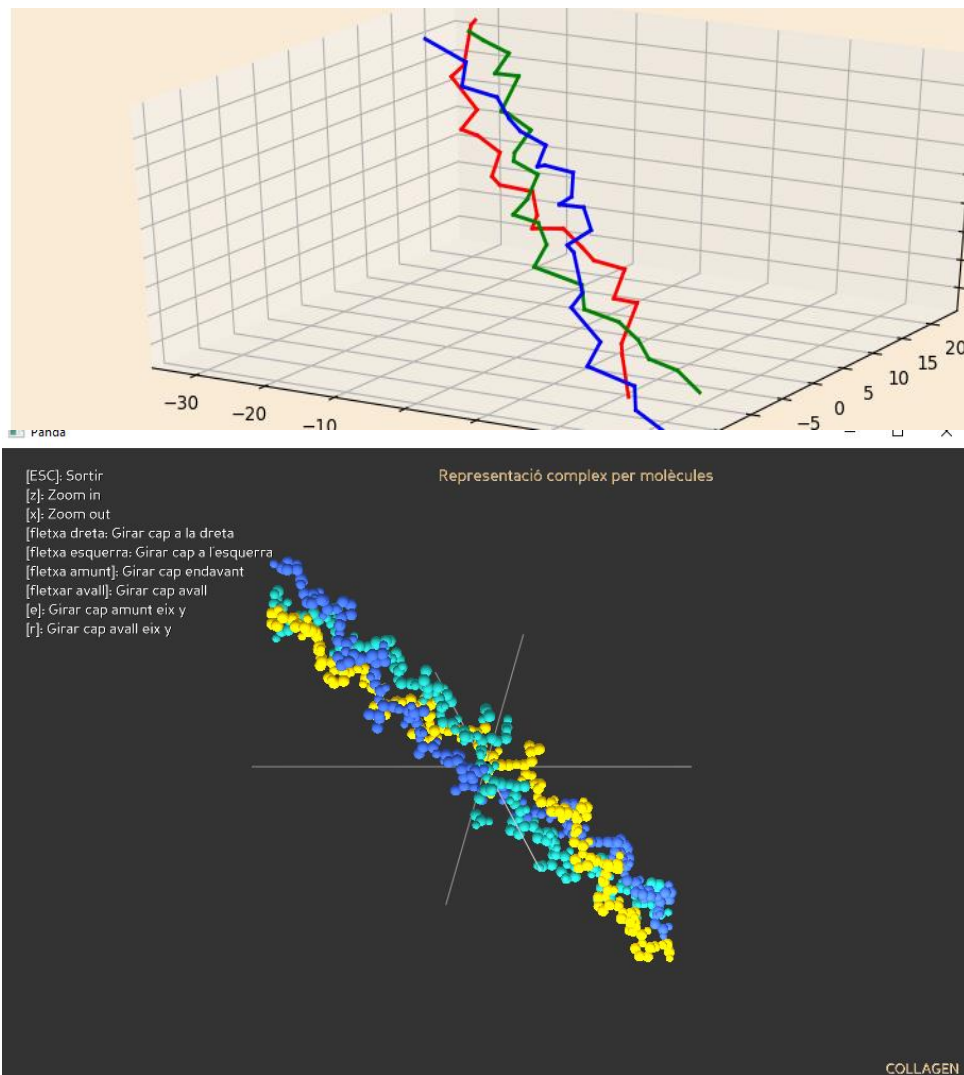


Figura 37.- Representació del colàgen per molècules en els dos sistemes

A l'observar les dues figures d'aquesta proteïna tant especial per la seva forma de fibres i no globular es pot veure que en el matplotlib si done una idea mol aproximada de la molècula.

Considero que he après molt sobre el llenguatge python i sobre la PPO, ja que només havia fet el curs del màster i fet tot amb notebook. Era la primera vegada que m'afrontava a un repte de fer un projecte sencer. Primer vaig tenir moltes dificultats que es van anar llimant, en gran part gràcies als consells del tutor i a l'aplicació pyCharm especialment a la utilització del debug al principi que em va ajudar a comprendre com tractava python les instruccions i les variables

Encara que sembli pretencios, estic molt satisfet dels resultats no esperava arribar tant lluny amb el panda3D que a primera vista veia una muntanya enorme. Al veure que me'n anava sortint l'autoexigencia es va incrementar i m'ha frustrat molt la impossibilitat d'aconseguir crear .egg tot i que ho vaig intentar a partir de Blender, haig de reconèixer que vaig fracassar.

Per sort el panda3D se'm va resistir menys de l'esperat i em va permetre avançar molt més del previst i per tant a pesar d'algunes circumstàncies adverses he aconseguit lograr l'objectiu que m'havia marcat (fins i tot més del que esperava)

El fet de que avancés més en el punt anterior em va permetre modificar l'objecte d'estudi que era senzillament estudiar la proteïna SERT o amb nomenclatura pdb la 5I6X.pdb . I vaig anar estudiant el funcionament amb diferents proteïnes que incloïen en el directori codi.dades_pdb 17 proteïnes. El fet de provar les diferents proteïnes han aparegut particularitats que no havia contemplat i m'han ajudat a millorar els mòduls.

Un tema que no m'havia plantejat però que ha sorgit al treballar amb les proteïnes ha sigut el dibuix de les estructures secundàries, especialment la helix α i les làmines β . Quan vaig consultar amb el tutor em va dir que els registres del .pdb "HELIX " i "SHEET " no s'utilitzaven i que es tenia que recórrer a la classe DSSP de Biopython, també he trobat la llibreria pytraj a http://amber-md.github.io/pytraj/release/1.0.7/dssp_analysis.html

Aquests dos darrers punts, construir models .egg i el bssp no puc dir que m'han quedat pendents, perquè no estaven previstos, però són temes que m'agradaria estudiar

8. GLOSARI

.pdb	Extensió dels fitxers més utilitzats per la representació de proteïnes del PDB
DSSP	Mòduls per tractar les estructures secundaries
Enllaç peptídic	Unió d'un grup amino d'un aminoàcid amb un grup carboxil d'un altre alliberant una molècula d'aigua
Estructura primària	Estructura primària d'una proteïna es la llista d'aminoàcids ordenats per formar una proteïna sense cap plegament no és funcional
Estructura secundària	Plegaments de l'estructura primària de varis tipus, els plegaments principals són l'hèlix α i les làmines β l'atracció es realitza amb ponts d'hidrogen
Estructura terciària	Resulta de la unió de varies estructures secundaries unides per ponts bisulfurs, pont d'hidrogen, forces de van der Waals, electrostàtics
Matplotlib	Llibreria gràfica clàssica per a python
Molècules	són les diferents parts de la proteïna formada per una cadena de peptídica i que li confereixen l'estructura quaternària
panda3D	Motor de videojocs dissenyat originàriament per Walt Disney es software lliure de 2008
PDB	Protein Data Bank
POO	Programació orientada a objectes
PyCharm	Editor IDE molt potent per treballar amb projectes de python https://www.jetbrains.com/pycharm/
residu	Resta de l'aminoàcid un cop unit a un altre aminoàcid mitjançant un enllaç peptídic i perduda la molècula d'aigua

9. BIBLIOGRAFIA

- Antao, T. (2015). *Bioinformatics with Python Cookbook*. Birmingham: Pakt Publishing Ltd.
- Arboledas, D. (2011). *Jerarquía estructural de las proteínas*. Alicante: Editorial club universitario.
- Bank, P. D. (2/10/2017). *Atomic Coordinate Entry Format Description Version 3.30*.
ftp://ftp.wwpdb.org/pub/pdb/doc/format_descriptions/Format_v33_Letter.pdf.
- Chang, J. C. (3/4/2018). *Biophyton Tutorial and Cookbook*.
<http://www.csc.kth.se/utbildning/kth/kurser/DD2397/appbio09/docs/BioPython.pdf>.
- Conlan, C. (2017). *The blender api 3D modelling and Add-on development*. Maryland: Apress.
- González, J. (2018). *Curso de biomoléculas*. Obtenido de Universidad del País Vasco - CSIC:
<http://www.ehu.eus/biomoleculas/index.htm>
- Gonzalez, E. (2013). Composición corporal: estudio y utilidad clínica. *Endocrinología y nutrición*, nº 60, 69-75.
- Jimenez, B. (10/10/2018). *Dockit. Visor de proteinaes emn Panda3D*.
<https://github.com/brianjimenez/dockit>.
- Kinsor, J. (1962). *Python for Bioinformatics*. London: Jones & Bartlett publishers.
- Lang, C. (2011). *Panda 3D 1,7 Game Developers CookBook*. Birmingham: Pakt Publishing Ltd.
- Mathews, D. (2011). *Panda 3D game Engine. Beginner`s Guide*. Birmingham: Pakt Publishing.
- Model, M. (2010). *Bioinformatics Programming using Python*. Sevastopol: O`Reilly Media Inc.
- Panda, 3. (29/9/2017). *Panda3D.- Main Page*.
https://www.panda3d.org/manual/index.php/Main_Page.
- Via, A., Rother, K., & Tramontano, A. (2014). *Managing you biological data with Python*. Boca Raton: CRC Press.

10. ANNEXS

Enllaços a visualitzadors moleculars i de proteïnes que es pot trobar a la pàgina web del PDB :

https://www.rcsb.org/pages/thirdparty/molecular_graphics

- [BioBlender](#)

Aplicació en codi obert que inclou funcions per transformar proteïnes i visualitzar potencials lipofílics i electroestàtics

- [BRAGI](#)

Un programa de visualització i modelatge de proteïnes

- [CCP4mg](#)

Crea belles imatges i pel·lícules amb qualitat de publicació. Els usuaris també poden superposar i analitzar estructures. El programa s'executa en plataformes Linux, MacOSX i Windows.

- [Chimera](#)

Sistema interactiu de modelatge molecular, gratuït per a acadèmics / sense fins de lucre; mostra múltiples alineaments de seqüències i estructures associades, de tipus àtom i identificació enllaç d'Hidrògens, trajectòries de dinàmica molecular (format amber), i ofereix interfície de lligament de detecció (moll), filtre per nombre / posició d'enllaços d'Hidrogen , i extensibilitat per crear encàrrec mòduls - per a Windows, Linux, Mac OS X, IRIX i Tru64 Unix

- [Cn3D](#)

Mostra simultàniament la estructura, seqüència i alineació, amb funcions d'edició d'anotació i alineació, per a usar amb estructures tridimensionals d'Entrez de NCBI. disponible per a Windows, Macintosh i Unix

- [CrystalMaker](#)

Un programa per construir, mostrar i manipular tot tipus d'estructures cristal·lines i moleculars.

- [ePMV](#)

El visualitzador molecular de Python incrustat (ePMV) és un plug-in de codi obert que executa programari de modelatge molecular directament dins de les aplicacions professionals d'animació 3D.

- [EzMol](#)

Una eina de visualització molecular basada en web fàcil d'usar, especialment dissenyada per a l'usuari ocasional que funciona amb els navegadors més comuns, pel que no hi ha necessitat de cap instal·lació o llicència. El model de visualització final pot descarregar-se per a la seva publicació o guardar-se per a la seva posterior ús.

- [Foldit](#)

És un joc d'ordinador de crowdsourcing basat en modelatge de proteïnes.

- [iMol](#)

El programa de gràfics iMol Open GL mostra molècules petites, grans i múltiples; mesura distàncies i angles, superposa estructures, calcula RMSD entre coordenades d'àtoms, alinea estructuralment cadenes i mostra trajectòries dinàmiques. Per a Mac OS X incl. 10.2

- [Jmol](#)

Jmol és un visor de molècules de codi obert per a estudiants, educadors i investigadors en química i bioquímica. Es multiplataforma i s'executa en sistemes Windows, Mac OS X i Linux / Unix.

- [Mage and Kinemages](#)

Visor molecular interactiu per a investigació i usos educatius. Font oberta i gratuïta per a Windows i Mac (OSX o PPC), Unix i Linux. Una versió de Java fa una visualització web 3-D sense complements.

- [Marvin](#)

Marvin és una col·lecció d'eines per dibuixar, visualitzar i caracteritzar estructures químiques, consultes, macromolècules i reaccions per a tots els sistemes operatius, pàgines web i aplicacions personalitzades.

- [MembraneEditor](#)

Genera de forma interactiva membranes basades en PDB heterogènies amb composicions lipídiques variables i col·locació semiautomàtica de proteïnes. Admet patilles de membrana i vesícules, micro-dominis i apilament de membranes monocapa i / o bicapa.

- [MOE \(Molecular Operating Environment\)](#)

Els gràfics de qualitat de publicació es poden generar en l'entorn operatiu molecular (MOE), que també ofereix una col·lecció d'innovadores formes per analitzar i mostrar sistemes de nivell molecular. MOE també conté una àmplia gamma d'aplicacions d'avantguarda per al modelatge de proteïnes i molècules petites, i el descobriment de fàrmacs. MOE és compatible amb els sistemes operatius Windows, Linux i Mac.

- [Molecule World](#)

Molecule World 2.1 és una aplicació d'iPad per visualitzar i manipular estructures químics i moleculars en 3D. Les estructures es poden descarregar i mostrar des de les bases de dades de l'estructura PubChem, PDB i NCBI juntament amb les seqüències de proteïnes i àcids nucleics. Les estructures es poden dibuixar com tubs, boles i barres, o modes d'emplenament d'espai. Les opcions de coloració inclouen molècules i residus, càrregues, hidrofobicitat, arc de Sant Martí. Les parts de les estructures es poden ocultar o mostrar amb colors barrejats i modes de dibuix.

- [Molecule World for iPhone](#)

Molecule World per iPhone es pot usar en l'iPhone o iPod Touch per visualitzar i manipular estructures químics i moleculars en 3D de les bases de dades de l'estructura PubChem, PDB o NCBI. Les opcions de dibuix inclouen modes de ball & stick, de superfície i ball & stick. Les opcions per

a acolorir inclouen arc de Sant Martí, residus, càrrega, hidrofobicitat i molècula. Les proteïnes, els àcids nucleics i els heterògens es poden mostrar en diferents maneres.

- [Molecule World DNA Binding Lab](#)

Una aplicació de l'iPad llista per a la classe per explorar les formes en què els productes químics i les proteïnes s'uneixen a l'ADN. El Laboratori de Vinculació de DNA utilitza el motor de renderitzat de Molecule World i les funcions de visualització per ressaltar diferents molècules i comprendre com es mantenen intactes. El laboratori d'unió d'ADN inclou instruccions, tres exemples i 40 incògnites que es poden assignar als estudiants. Les capacitats per compartir fotos permeten als estudiants compartir el seu treball amb els mestres per ajudar amb l'avaluació.

- [Molecules](#)

Una aplicació d'iPhone per estructures PDB

- [MolScript](#)

Un programa per mostrar estructures en formats detallats i esquema

- [MolviZ.org](#)

Tutorial de visualització interactius gratuïts

- [MVM](#)

Programa de visualització molecular i GUI de ZMM. MVM és un visor molecular gratuït que es pot usar per mostrar proteïnes, àcids nucleics, oligosacàrids, petits i macromolècules. Té una interfície intuïtiva. A més de ser un visor molecular, és la interfície d'usuari d'un motor de molta força mecànica (ZMM).

- [PMV \(Python Molecular Viewer\)](#)

Un entorn interactiu de visualització i modelatge molecular per a la manipulació i visualització de múltiples molècules.

- [PocketMol](#)

Programa PocketMol per veure i manipular arxius PDB en un PocketPC

- **POLYVIEW**

- [POLYVIEW-2D](#)

Anotació de la estructura de proteïna utilitzant perfils de seqüència

- [POLYVIEW-3D](#)

Anotació versàtil y visualització de alta qualitat d'estructures macromoleculares

- [POLYVIEW-MM](#)

Anàlisis y visualització de moviments macromoleculares

- [ProteinScope](#)

Visor gratuït per mostrar i manipular arxius PDB i crear animacions i diapositives de proteïnes per a Windows. Comandes en línia d'impressions de proteïnes 3D en diversos esquemes de color. (Antao, 2015)

- [Prosat](#)

Representar anotacions de seqüències de proteïnes en una estructura de proteïnes i visualitzar-les simultàniament amb l'estructura.

- [PyMOL](#)

Un sistema de gràfics moleculars lliures i de codi obert per a visualització, animació, edició i imatges amb qualitat de publicació. PyMOL és programable i es pot estendre utilitzant el llenguatge Python. Admet Windows, Mac OSX, Unix i Linux

- [QuteMol](#)

Un sistema de visualització molecular de gran qualitat, interactiu i de codi obert (GPL). QuteMol explota les capacitats actuals de GPU a través d'ombres OpenGL per oferir una varietat d'efectes visuals innovadors.

- [RasMol](#)

Un sistema de visualització gratuïta per a fitxers de coordenades PDB que es executa en sistemes Mac (PPC), Windows, Unix i Linux. Les versions de codi obert també estan disponibles.

- [Raster3D](#)

Un conjunt d'eines per generar imatges ràster d'alta qualitat de proteïnes o altres molècules. Freeware per Mac OSX, Windows, Unix i Linux

- [RasTop \(v. 2.0\)](#)

Una interfície gràfica d'usuari gratuïta per al programari de visualització molecular RasMol (v. 2.7.2.1), disponible per a Windows i Linux

- [Ribbons](#)

Un programa per a la il·lustració molecular i anàlisi d'errors, per Mac OSX, Windows, Unix i Linux

- [RCSB MBT Viewers](#)

El kit d'eines MBT és un marc que permet crear diverses viseres. Es fa servir per a 4 visors diferents a la pàgina web de RCSB PDB.

- [RmscopII](#)

Un script Tcl / Tk responsable de redirigir arxius PDB o scripts RasMol a múltiples sessions RasMol; es pot usar com una aplicació de navegador web o com un programa independent per a Mac (OSX o PPC), Windows o Unix

- [Schrödinger Product Suites](#)

Les ofertes de productes complets de Schrödinger Product Suites van des de programes generals de modelatge molecular fins a un conjunt complet de programari de disseny de fàrmacs, així com un paquet de avantguarda per a la investigació de materials. Tots els productes es executen amb el Mestre, una interfície unificada per a tot el programari Schrödinger, que està disponible per a Mac, Windows i Linux.

- [SPADE](#)

L'entorn de desenvolupament d'aplicacions de Estructura Proteòmica (SPADE) proporciona eines comunitàries per al desenvolupament i la implementació d'equips essencials de seqüència i estructura. Inclou un conjunt de proves químics per a suportar la verificació experimental de

models estructurals predits. Escrit en Python amb eines de scripts disponibles. Es executa en Windows, Linux i Mac.

- [STRAP](#)

Alinea les proteïnes per seqüència i estructura 3D.

- [Swiss PDB viewer](#)

Un programa de gràfics 3D i modelat molecular per a l'anàlisi simultània de múltiples models i per a la construcció de models en mapes de densitat electrònica. El programari està disponible per Mac (OSX o PPC), Windows, Linux o SGI

- [UGENE](#)

Una eina gratuïta i de codi obert amb suport de visualització de format PDB escrit en codi C ++ de ràpid memòria eficient. Admet Windows, Mac OSX, Unix i Linux.

- [VMD](#)

VMD (Visual Molecular Dynamics) s'executa en moltes plataformes, incloent MacOS X, i diverses versions d'Unix i Windows. VMD proporciona característiques de visualització, anàlisi i seqüències de comandaments Tcl/Python, i recentment ha afegit característiques d'exploració de seqüència i de representació volumètrica. VMD es distribueix de forma gratuïta.

- [YASARA](#)

Un complet programa de modelatge i gràfics moleculars, que inclou simulacions interactives de dinàmica molecular, determinació d'estructura, anàlisi i predicció, acoblament, pel·lícules i eLearning per a Windows, Linux i MacOSX.

- [Zeus](#)