Citlali Gil Aguillon
Jessica Danielly Medina
Analí Migueles Lozano

Tarea algoritmos3D-4

OBJETIVO: Aprender a modelar y evaluar un modelo 3D de una proteína.

1) Elige una secuencia S de la superfamilia que elegiste para la tarea 3.

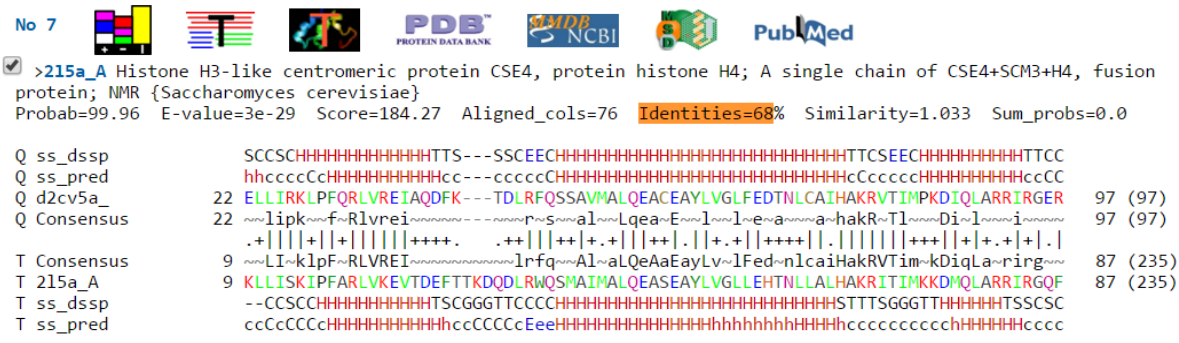
Se eligió: Species: Human (Homo sapiens), H2A.a [TaxId: 9606], cuya entrada de PDB es 2cv5

2) Usando HHpred (http://toolkit.tuebingen.mpg.de/hhpred) selecciona al menos una estructura molde o template que puedas usar para modelar S, asegurándote que tiene menos del 90% de identidad si fuera posible.

No	Hit		Prob	E-value	P-value	Score	SS	Cols	Query HMM	Template HMM
1	2yfv_A	Histone H3-like centrom	100.0	2.6E-45	7.1E-50	239.7	7.7	94	1-94	4-100 (100)
2	3nqu_A	Histone H3-like centrom	100.0	1.5E-43	4E-48	243.1	8.8	97	1-97	38-136 (140)
3	3r45_A	Histone H3-like centrom	100.0	3.7E-43	1E-47	244.3	7.2	95	1-95	54-150 (156)
4	1tzy_C	Histone H3; histone-fol	100.0	4E-40	1.1E-44	225.2	9.7	97	1-97	39-135 (136)
5	2hue_B	Histone H3; mini beta s	100.0	6.6E-33	1.8E-37	173.6	8.6	76	22-97	1-76 (77)
6	3nqj_A	Histone H3-like centrom	100.0	1.7E-32	4.6E-37	173.4	8.2	76	22-97	1-78 (82)
7	2l5a_A	Histone H3-like centrom	100.0	3E-29	8.2E-34	184.3	7.6	76	22-97	9-87 (235)
8	4xyi_B	Histone H4; centromere,	99.9	7.8E-27	2.2E-31	151.2	11.0	89	1-97	9-97 (103)
9	2ly8_A	Budding yeast chaperone	99.9	7.5E-27	2.1E-31	156.8	8.2	72	24-95	1-113 (121)
10	4jjn_B	Histone H4; BAH domain,	99.9	5.4E-28	1.5E-32	156.7	1.4	87	1-95	8-94 (102)
11	2yfw_B	Histone H4, H4; cell cy	99.9	7E-26	1.9E-30	146.8	5.8	87	1-95	9-95 (103)
12	1taf_B	TFIID TBP associated fa	99.9	4.7E-24	1.3E-28	130.3	6.3	70	19-93	1-70 (70)
13	3b0c_T	CENP-T, centromere prot	99.8	6.4E-20	1.8E-24	120.1	8.3	72	19-95	2-73 (111)
14	4csr_A	Nuclear transcription f	99.8	1.2E-19	3.3E-24	115.3	7.0	73	18-95	3-77 (94)
15	2hue_C	Histone H4; mini beta s	99.8	1.5E-19	4.2E-24	113.5	4.7	72	19-95	5-76 (84)
16	1tzy_D	Histone H4-VI; histone-	99.8	3.3E-19	9.2E-24	115.8	5.6	76	15-95	20-95 (103)
17	1b67_A	Protein (histone HMFA);	99.8	7.6E-19	2.1E-23	105.8	5.9	63	25-92	3-65 (68)
18	1ku5_A	HPHA, archaeal histon;	99.8	8.6E-19	2.4E-23	105.9	6.0	63	25-92	7-69 (70)
60	4zri_C	Serine/threonine-protei	89.7	0.22	6.1E-06	25.8	1.7	15	2-16	1-15 (32)
61	4wv4_A	Transcription initiatio	87.9	1.5	4.2E-05	28.3	5.2	81	11-96	6-99 (102)
62	4zrk_E	Serine/threonine-protei	85.4	0.7	1.9E-05	23.9	2.0	15	2-16	1-15 (32)
93	1um8_A	ATP-dependent CLP prote	39.4	1E+02	0.0029	21.9	6.0	59	30-88	289-359 (376)
94	1ixz_A	ATP-dependent metallopr	37.1	47	0.0013	22.4	3.7	29	63-91	225-253 (254)
95	3k6q_A	Putative ligand binding	36.9	70	0.0019	21.1	4.4	33	49-81	81-113 (139)
96	1w5s_A	Origin recognition comp	36.5	1.4E+02	0.0039	21.7	8.2	63	30-92	220-291 (412)
97	1njg_A	DNA polymerase III subu	34.6	1.1E+02	0.0031	19.9	6.3	58	30-90	190-247 (250)
98	1iy2_A	ATP-dependent metallopr	32.7	70	0.0019	21.8	4.1	28	63-90	249-276 (278)
99	4fdg_B	Minichromosome maintena	30.7	30	0.00083	28.3	2.1	50	44-93	531-598 (693)

Se eligieron las secuencias con 68% y 65% de identidad, las cuales eran las que presentaban los porcentajes de identidad más altos.

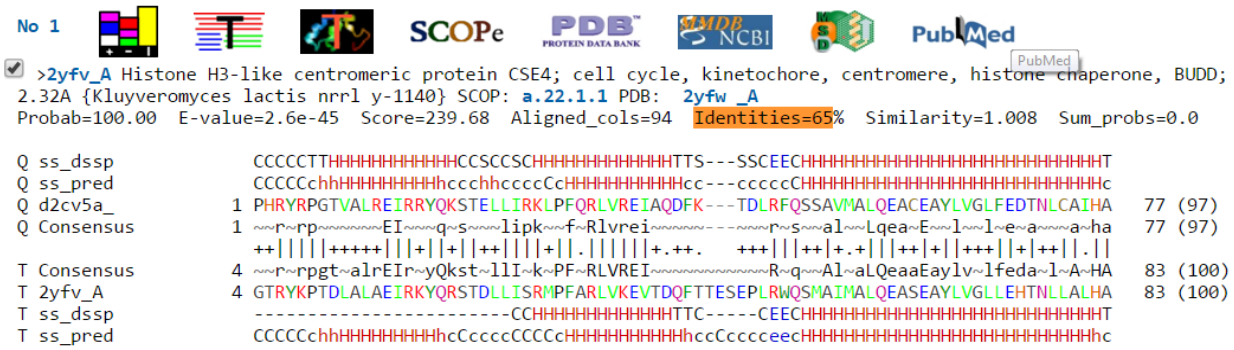
Identidad=68%



Este hit es de una proteína nuclear. Es la base estructural de una variante de la histona H3 que tiene un reconocimiento específico hacia el centrómero.

Identidad=65%

Citlali Gil Aguillon
Jessica Danielly Medina
Analí Migueles Lozano



3) De acuerdo con el ejemplo de http://eead-csic-compbio.github.io/bioinformatica_estructural/node34.html y la documentación de MODELLER construye dos modelos M1 y M2 de S y comprueba su estima de calidad con DOPE

Sabemos que MODELLER es una herramienta para hacer modelado comparativo y/o de homología. Utilizando MODELLER en línea de comando se pudieron obtener los modelos M1 y M2 y luego se tuvo que comprobar su calidad con DOPE. DOPE es un potencial estadístico que se usa para evaluar modelos de las estructuras terciarias de proteínas.
Cuando se corrió la prueba DOPE, éste te regresa molpdfs; los valores más pequeños son los mejores y dichos modelos son considerados como los más óptimos.
En sí, DOPE está diseñado para seleccionar la mejor estructura de una colección de modelos que MODELLER construyó.

Modelo 1 (M1)
Para el modelo 1, el valor obtenido de DOPE es de -8717.161133, lo que quiere decir que es confiable.

```
                University of California, San Francisco, USA
                   Rockefeller University, New York, USA
                      Harvard University, Cambridge, USA
                   Imperial Cancer Research Fund, London, UK
              Birkbeck College, University of London, London, UK


Kind, OS, HostName, Kernel, Processor: 4, Windows Vista build 7601 Service Pack 1, ANALÍ-PC, SMP, unknown
Date and time of compilation         : 2016/01/07 09:07:44
MODELLER executable type             : x86_64-w64
Job starting time (YY/MM/DD HH:MM:SS): 2016/03/06 22:15:52

read_to_681_> topology.submodel read from topology file:        3

getf_______w> RTF restraint not found in the atoms list:
              residue type, indices:    13     1
              atom names           : N     -C     CA     CD
              atom indices         :     1      0      2      3
>> Model assessment by DOPE potential
iatmcls_286W> MODEL atom not classified:  ARG:OXT  ARG
preppdf_453W> No fixed restraints selected; there may be some dynamic ones.
preppdf_454W> Restraints file was probably not read; use restraints.append().


>> ENERGY; Differences between the model's features and restraints:
Number of all residues in MODEL                  :       97
Number of all, selected real atoms               :      802     802
Number of all, selected pseudo atoms             :        0       0
Number of all static, selected restraints        :        0       0
COVALENT_CYS                                     :        F
NONBONDED_SEL_ATOMS                              :        1
Number of non-bonded pairs (excluding 1-2,1-3,1-4):    82626
Dynamic pairs routine                            : 1, NATM x NATM double loop
Atomic shift for contacts update (UPDATE_DYNAMIC) :    0.390
LENNARD_JONES_SWITCH                             :    6.500   7.500
COULOMB_JONES_SWITCH                             :    6.500   7.500
RESIDUE_SPAN_RANGE                               :        1    9999
NLOGN_USE                                        :       15
CONTACT_SHELL                                    :   15.000
DYNAMIC_PAIRS,_SPHERE,_COULOMB,_LENNARD,_MODELLER :        T       F       F       F       T
SPHERE_STDV                                      :    0.050
RADII_FACTOR                                     :    0.820
Current energy                                   :       -8717.1611



<< end of ENERGY.
DOPE score               : -8717.161133
Total CPU time [seconds]                         :       3.87
```

Citlali Gil Aguillon
Jessica Danielly Medina
Analí Migueles Lozano

Modelo 2 (M2)

Para el modelo 2, el valor obtenido de DOPE es de -19281.568359, lo que quiere decir que es un modelo confiable.

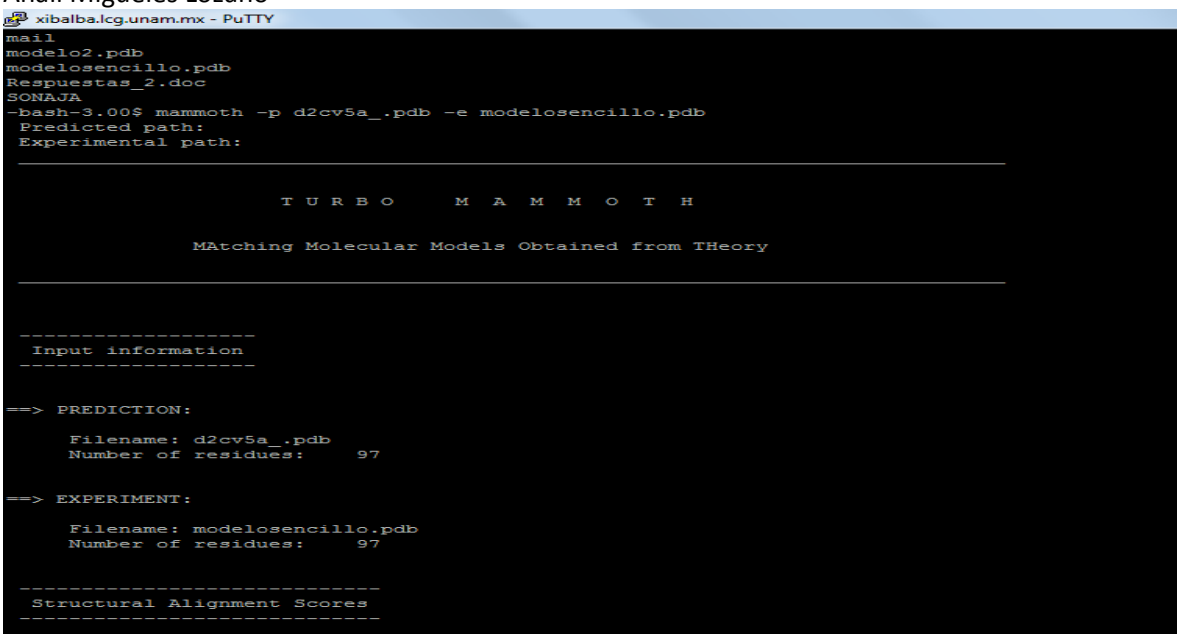
```
>> Model assessment by DOPE potential
iatmcls_286W> MODEL atom not classified:  GLY:OXT  GLY
preppdf_453W> No fixed restraints selected; there may be some dynamic ones.
preppdf_454W> Restraints file was probably not read; use restraints.append().


>> ENERGY; Differences between the model's features and restraints:
Number of all residues in MODEL                  :      225
Number of all, selected real atoms               :     1803    1803
Number of all, selected pseudo atoms             :        0       0
Number of all static, selected restraints        :        0       0
COVALENT_CYS                                     :        F
NONBONDED_SEL_ATOMS                              :        1
Number of non-bonded pairs (excluding 1-2,1-3,1-4):   312900
Dynamic pairs routine                            : 1, NATM x NATM double loop
Atomic shift for contacts update (UPDATE_DYNAMIC) :    0.390
LENNARD_JONES_SWITCH                             :    6.500   7.500
COULOMB_JONES_SWITCH                             :    6.500   7.500
RESIDUE_SPAN_RANGE                               :        1    9999
NLOGN_USE                                        :       15
CONTACT_SHELL                                    :   15.000
DYNAMIC_PAIRS,_SPHERE,_COULOMB,_LENNARD,_MODELLER :        T       F       F       F       T
SPHERE_STDV                                      :    0.050
RADII_FACTOR                                     :    0.820
Current energy                                   :      -19281.5684




<< end of ENERGY.
DOPE score               : -19281.568359
Total CPU time [seconds]                         :       4.63
```

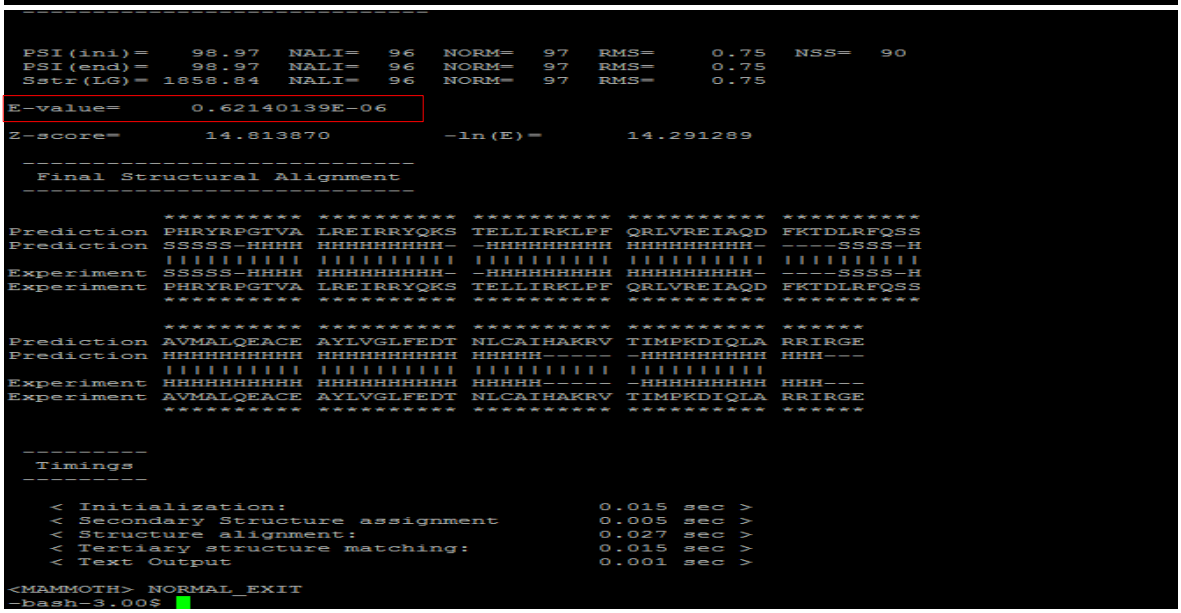
4) Evalúa la calidad de los modelos M obtenidos comparándolos con la estructura conocida, que descargaste de SCOP en la tarea 3. Para ello puedes usar MAMMOTH. En tu informe por favor indica el alineamiento obtenido, el RMSD y al menos una imagen de su superposición para brevemente comentar las diferencias que observas entre cada modelo y la estructura experimental.

MODELO 1

- Alineamiento en MAMMOTH

Citlali Gil Aguillon
Jessica Danielly Medina
Analí Migueles Lozano





- RMSD (utilizando el script proporcionado en : http://eead-csic-compbio.github.io/bioinformatica_estructural/node31.html)

total residuos: pdb1 = 97 pdb2 = 97

total residuos alineados = 96

coordenadas originales = original.pdb

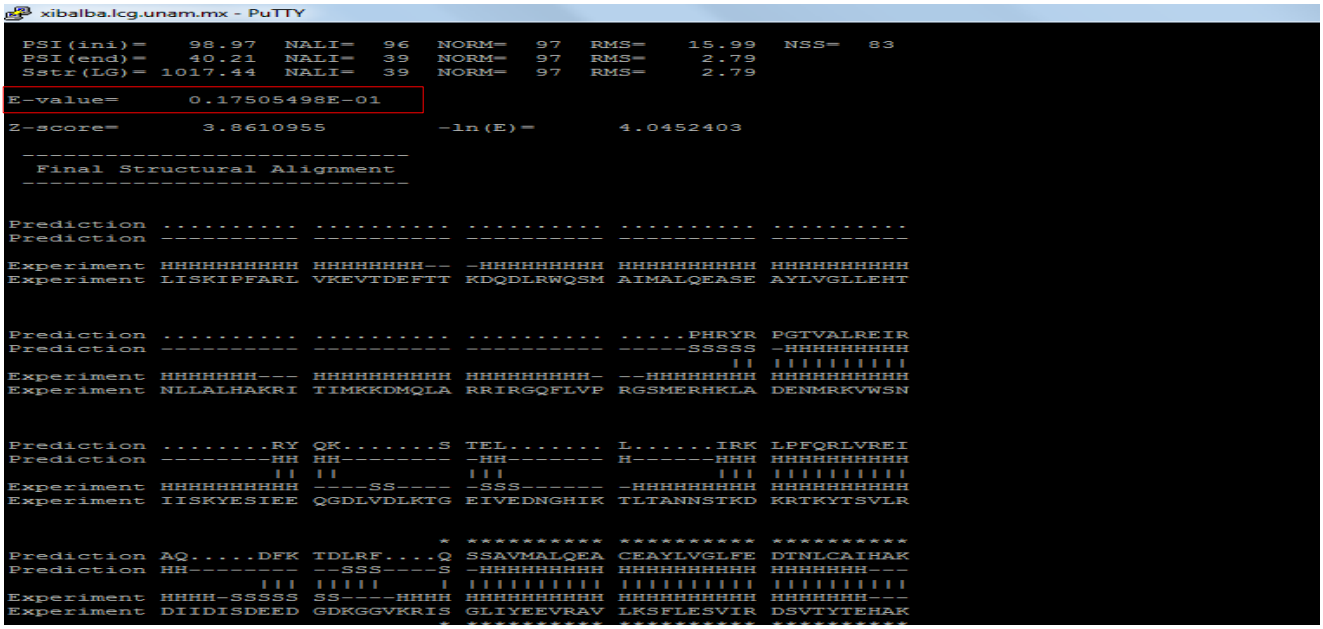
superposicion optima:

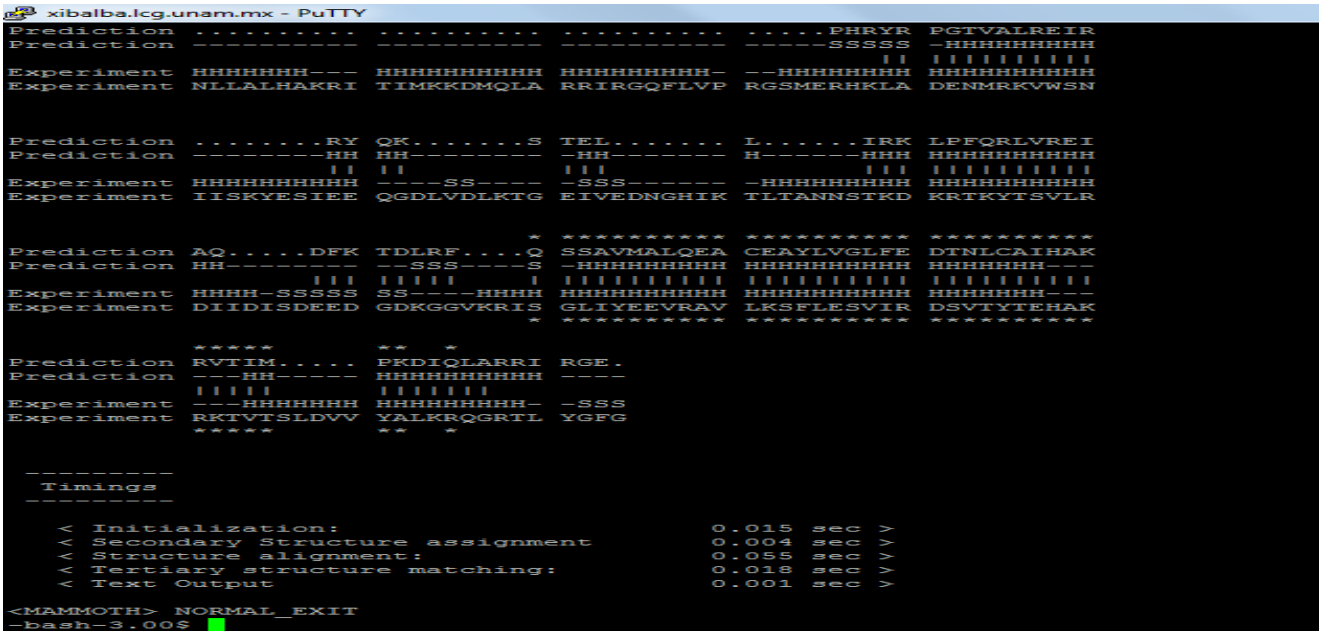
archivo PDB = align_fit.pdb

RMSD = 0.75 Angstrom

Citlali Gil Aguillon
Jessica Danielly Medina
Analí Migueles Lozano
MODELO 2

Alineamiento en MAMMOTH





- RMSD (el script proporcionado en : http://eead-csic-compbio.github.io/bioinformatica_estructural/node31.html)

total residuos: pdb1 = 225 pdb2 = 97

total residuos alineados = 96

coordenadas originales = original.pdb

superposicion optima:

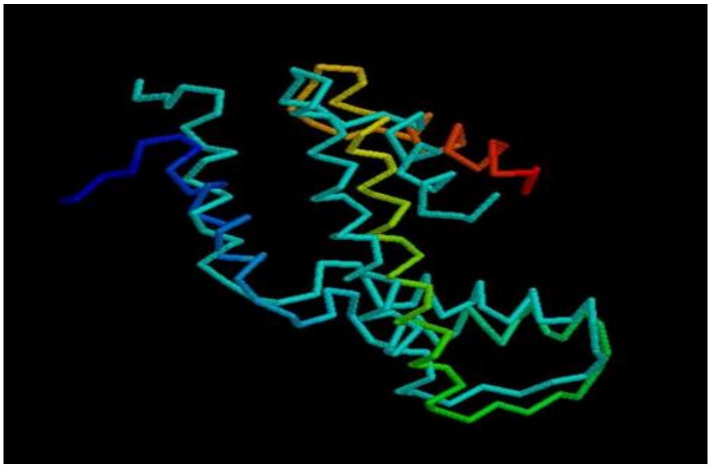
Citlali Gil Aguillon
Jessica Danielly Medina
Analí Migueles Lozano
archivo PDB = align_fit.pdb

RMSD = 15.99 Angstrom

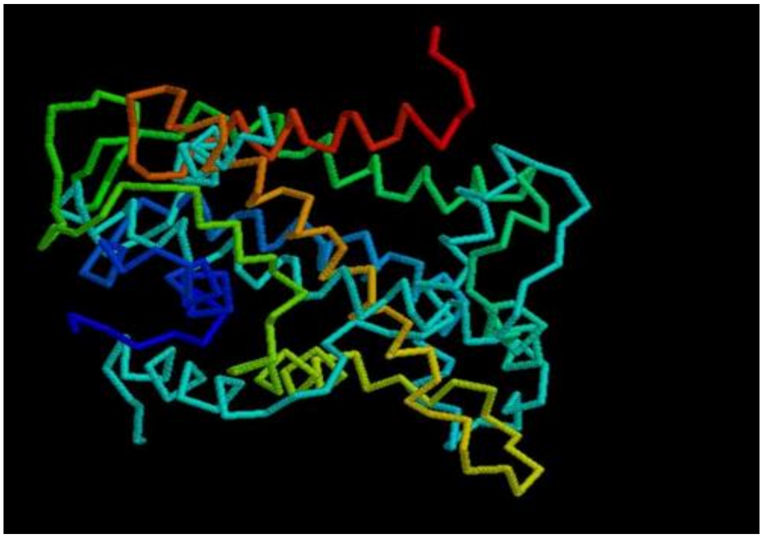
::

IMÁGENES

Query-modelo 1



Query-modelo 2



Query –Modelo1-Modelo2

Citlali Gil Aguillon
Jessica Danielly Medina
Analí Migueles Lozano

CONCLUSIÓN

Observando los valores de RMSD y los de e-value podemos concluir que el modelo 1 es el de mejor calidad porque su valor de RMSD es el más pequeño y su e-value definitivamente mejor que el del modelo 2. Como podemos observar en las imágenes, el modelo 1 sobrepuesta con la proteína query se puede apreciar una gran similitud. Dicho modelo mostraba un DOPE score bastante aceptable. Su RMSD, que es una medida para decir que tan alejados son las proteínas entre sí, era igualmente prometedor (era pequeño), su e-value no eran tan negativo, pero estaba a la 1x10ˆ-6.

Si lo comparamos con el modelo 2, vemos diferencias abismales. Si bien el valor de DOPE score era prometedor, su RMSD era muy grande (15.99) y su e-value era pésimo. Observando la imagen se ve claramente que no tienen casi nada de similitud experimental. Esto nos demuestra que se necesitan de más pruebas –tanto estadísticas como experimentales y bioinformáticas- para poder evaluar y obtener modelos que se asemejen más a la realidad. Si bien lo anterior sigue siendo un gran obstáculo en el ámbito de la predicción estructural de proteínas, rescatar los puntos débiles de los modelos predictivos nos hace acercarnos cada vez más a métodos mas acertados.