Practica4

Carmina Barberena Jonas

Tarea algoritmos3D-4

La tarea 4 consiste en aprender a modelar y evaluar un modelo 3D de una proteína:

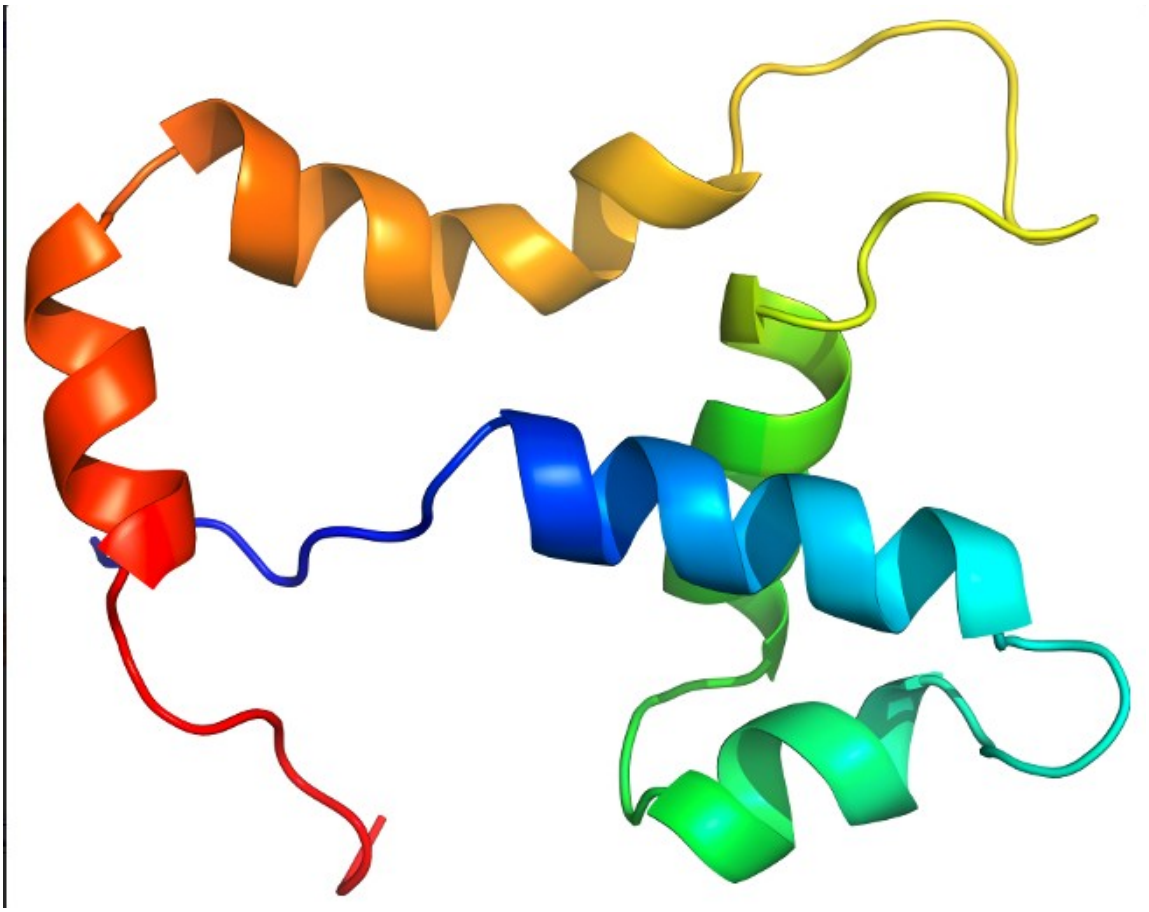
1 Elige una secuencia S de la superfamilia que elegiste para la tarea 3.

Class a: All alpha proteins.

Superfamily a.4.1: Homeodomain-like.

Species :Corynebacterium glutamicum

```
>d2jn6a1 a.4.1.19 (A:1-89) Uncharacterized protein Cgl2762
{Corynebacterium glutamicum [TaxId: 1718]}
mptktyseefkrdavalyensdgaslqqiandlginrvtlknwiikygsnhnvqgttpsa
avseaeqirqlkkenalqrartrhpaesc
```



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.

3hug_A RNA polymerase sigma factor Identidad=33%

```
>3hug_A RNA polymerase sigma factor; ECF sigma factor, zinc binding anti-sigma factor, oxidative transcription
regulation; 2.35A {Mycobacterium tuberculosis}
Probab=97.05  E-value=0.00043  Score=36.79  Aligned_cols=43  Identities=33%  Similarity=0.268  Sum_probs=0.0

Q ss_dssp             CCCHHHHHHHHHHTTGGGSCHHHHHHHHTSCHHHHHHHHHHHCC
Q ss_pred             cCCHHHHHHHHHHHHhCCCCHHHHHHHHhCCCHHHHHHHHHHHhh
Q d2jn6a1          5 TYSEEFKRDAVALYENSDGASLQQIANDLGINRVTLKNWIIKYGS   49 (89)
Q Consensus        5 ~~~~e~k~~~~~~~~~~~~s~~~ia~~~~vs~~ti~~w~~~~~~   49 (89)
                     .+++..+.-+...+.+  |+|..+||..+|+|.+||+.++.+...
T Consensus       37 ~L~~~~r~il~l~~~~--g~s~~eIA~~l~is~~tv~~~~~ra~~   79 (92)
T 3hug_A          37 QLSAEHRAVIQRSYYR--GWSTAQIATDLGIAEGTVKSRLHYAVR   79 (92)
T ss_dssp             TSCHHHHHHHCCCCCS--CCCHHHHHHHHTSCHHHHHHHHHHHHH
T ss_pred             hCCHHHHHHHHHHHHc--CCCHHHHHHHHCcCHHHHHHHHHHHHH
```

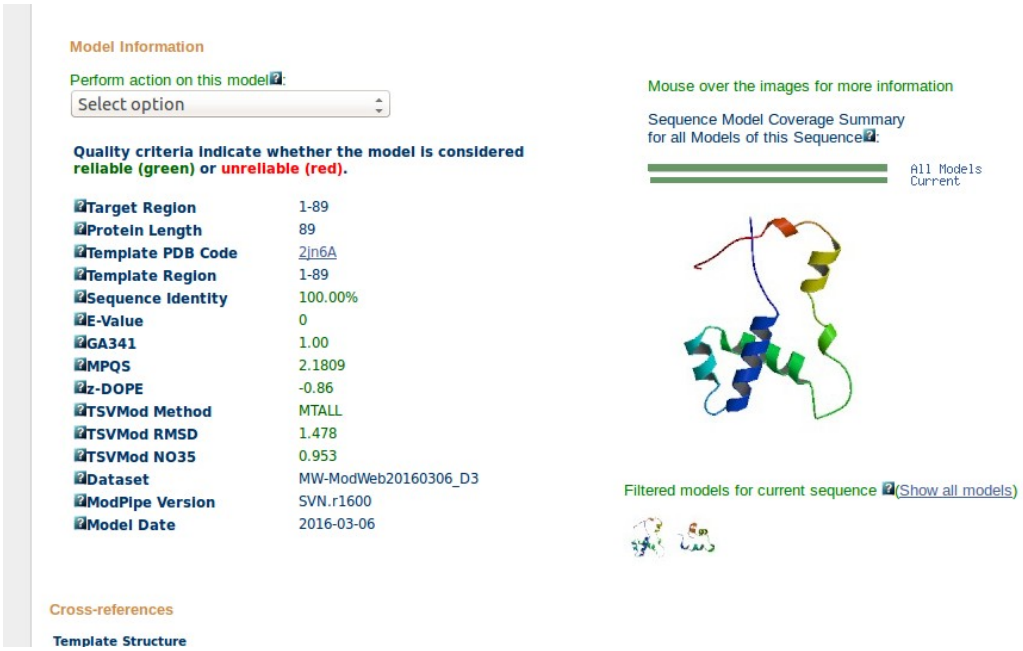
2ao9_A Phage protein Identidad=34%

```
>2ao9_A Phage protein; structural genomics, nine-fold NCS., PSI, protein structure initiative, midwest center for
structural genomics, MCSG, U function; 1.90A {Bacillus cereus} SCOP: a.4.1.17
Probab=98.82  E-value=2.7e-09  Score=63.30  Aligned_cols=47  Identities=34%  Similarity=0.470  Sum_probs=0.0

Q ss_dssp             CCCCC-----CCHHHHHHHHHHTTGG--------GSCHHHHHHHTSCHHHHHHHHHHC
Q ss_pred             CCCCc-----CCHHHHHHHHHHHHhC--------CCCHHHHHHhCCCHHHHHHHHHHh
Q d2jn6a1           1 MPTKT-----YSEEFKRDAVALYENSD--------GASLQQIANDLGINRVTLKNWIIKYG   48 (89)
Q Consensus         1 m~~~~-------~~e~k~~~~~~~~~~~~~~~~~~~~s~~~ia~~~~vs~~ti~~w~~~~~   48 (89)
                      |++++     |+++++..++.++.+ +        |.|..+||+.+|||++|+|+|++.+.
T Consensus        14 M~~~~~r~k~s~e~~~a~~l~~~.~~~~~g~~g~s~~eiA~~lgVs~~Tlyrw~~~~~   73 (155)
T 2ao9_A           14 MAKLDELKQKLTAKQIQAAYLLVEN-ELMESNNEEKRTQDEMANELGINRTTLWEWRTKNQ   73 (155)
T ss_dssp             HHHHHHHHTTSCHHHHHHHHHHHHH-HHCC---CCCCCHHHHHHHHTCCHHHHHHHHHCH
T ss_pred             ccccccCCCCCCHHHHHHHHHHHHc-cccccccccCCCHHHHHHHCcCHHHHHHHHHHcc
```

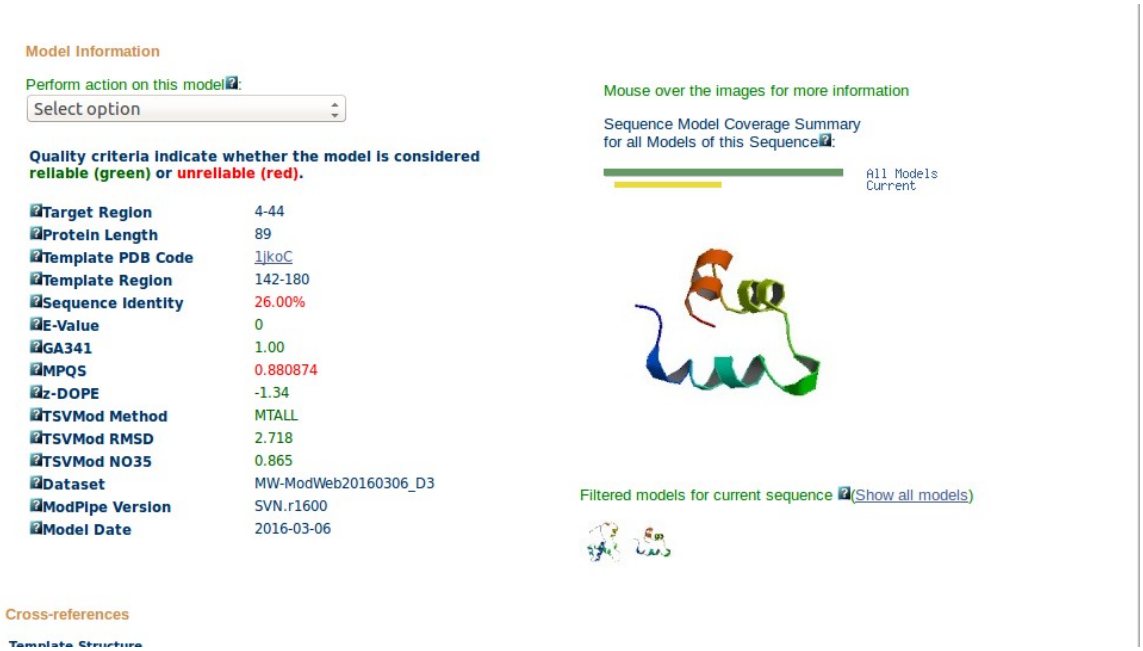
3 De acuerdo con el ejemplo de http://eead-csic-compbio.github.io/bioinformatica_estructural/node34.html y la documentación de MODELLER construye dos modelos M1 y M2 de S y comprueba su estima de calidad con DOPE.

M1



Este modelo que contiene mejores valores

M2



Dope

```
# Example for: selection.assess_dope()

from modeller import *
from modeller.scripts import complete_pdb

env = environ()
env.libs.topology.read(file='$(LIB)/top_heav.lib')
env.libs.parameters.read(file='$(LIB)/par.lib')

# Read a model previously generated by Modeller's automodel class
mdl = complete_pdb(env, 'modelosencillo.pdb')

# Select all atoms in the first chain
atmsel = selection(mdl.chains[0])

score = atmsel.assess_dope()
```

Resultados Modelo 1

ENERGY; Differences between the model's features and restraints: Number of all residues in MODEL : 41 Number of all, selected real atoms : 330 330 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): 34729 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 : -3465.5688

<< end of ENERGY.

Resultados Modelo 2

>> ENERGY; Differences between the model's features and restraints: Number of all residues in MODEL : 89 Number of all, selected real atoms : 700 700 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): 80660 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 : -6731.4219