Bioinformática estructural

Tarea 4-Algoritmos 3D

Carina Paola Cornejo Páramo

Claudia Saraí Reyes Ávila

Marzo-2016

Introducción

La estructura terciaria de una proteína está determinada por su secuencia, sin embargo el número de conformaciones posibles que puede adquirir son demasiadas, ya que se consideran los enlaces del backbone de la proteína así como las interacciones que puede establecer(puentes de hidrógeno, interacciones electrostáticas, de Van Der-Waals, hidrofóbicas etc.) Dado el gran número de conformaciones no es fácil predecir la conformación que es funcional, por ello para obtener una conformación aproximada de una proteína se usan moldes o templetes que son proteínas relacionadas en secuencia con nuestra proteína de interés, debido a que son homólogas. En el presente trabajo seleccionamos una proteína y evaluamos su estructura tridimensional usando templetes como guía.

Desarrollo

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

En la tarea 3 seleccionamos 3 proteínas de la superfamilia de los citocromos P450, para esta tarea seleccionamos una de éstas, la proteína d2j0da1 de $Homo\ sapiens$.

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.

En la siguiente imagen se muestra el templete que seleccionamos, tiene un 22% de identidad.

```
PDB
    No 2
                                                                                                                                                                                                                          Pub Med
                                                                                                                                                         S NCBI
>3tbg_A Cytochrome P450 2D6; monooxygenase, thioridazine, oxidoreductase; HET: RTZ HEM; 2.10A {Homo sapiens} PDB: 3qm4 _A* 3tda _A* 4wnt _A* 4wnu _A* 4wnv _A* 4wnw _A* 4xry _A* 4xrz _A* 2f9q _A* Probab=100.00 E-value=3.2e-73 Score=556.26 Aligned_cols=437 Identities=22% Similarity=0.337 Sum_probs=0.0
                                                                       HHHTCCCCCCBTTTBTGGGGG-GHHHHHHHHHHHHHHHHHTTCEEEEETTEEEEECCCHHHHHCCCCCCSTTTTCCCCCCCCCG
    Q ss_dssp
                                                                       hhcccccccchhih-ccccchhihh-cccccchhihh-cccccchhihh
    Q ss_pred
                                                                5 KKLGIPGPTPLPFLGNILSYH-KGFCMFDMECHKKYGKVWGFYDGQQPVLAITDPDMIKTVLVKECYSVFTNRRPFGPVG
    Q d2j0da1
                                                                                                                                                                                                                                                                                                                                  83 (445)
                                                                5 nannapppaapaaganaan-anaanaanaanyyyainaanaanaanyayaapaanainaanaana
    Q Consensus
                                                                                                                                                                                                                                                                                                                                   83 (445)
                                                                       8 \ \text{numpPGP} \\ \text{numpPGP} \\ \text{numpnamenum} \\ \text{lowyGNV} \\ \text{numpramenum} \\ \text{lowillamenum} \\ \text{lowillamenum} \\ \text{lower} \\ \text{lowillamenum} \\ \text{lower} \\
                                                                                                                                                                                                                                                                                                                                   86 (479)
    T Consensus
                                                                8 KGKLPPGPLPLPGLGNLLHVDFQNTPYCFDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTH-GEDTADRPPVPITQ
    T 3tbg_A
                                                                                                                                                                                                                                                                                                                                   86 (479)
     T ss_dssp
                                                                        -CCCCCCSCCBTTTBTGGGCCTTSHHHHHHHHHHHHHHCSEEEEEEETTEEEEEEEHHHHHHCCCCCT-GGGSCBCCCCGGGG
    T ss_pred
                                                                       Q ss_dssp
                                                                     Q ss_pred
     Q d2j0da1
                                                             84 FM-----KSA-ISIAEDEEWKRLRSLLSPTFTSGK--LKEMVPIIAQYGDVLVRNLRREAETGKPVTLKDVFGAYSMDVI 155 (445)
                                                            84 na - - - - nave lamos garynan RKanana fanna - lamana jananan lama lamanan volunan navana ii
    Q Consensus
                                                                                                                                                                                                                                                                                                                             155 (445)
                                                                     T Consensus
                                                             87 nunnnunnun lann gar Wann Rrannur fannunnun innunn lan lan - angan dannunnun wil
                                                                                                                                                                                                                                                                                                                               164 (479)
                                                             87 ILGFGPRSQGVFLARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAACLCAAFAN--HSGRPFRPNGLLDKAVSNVI
     T 3tbg_A
                                                                                                                                                                                                                                                                                                                               164 (479)
                                                                       GGTCBTTBCCSTTCCSSHAHAHAHAHAHAHAHAHTTSTTCHHAHAHAHAHAHAHAHAHAHAT--TTTCCBCTHAHAHAHAHAHAH
    T ss_dssp
                                                                       T ss_pred
    Q ss_dssp
                                                                       ННННСССССССССССССhhhhhннннн------ннннннhhhhhhнннн-----
    Q ss_pred
                                                          156 TSTSFGVNIDSNPQDPFVENTKKLLRFF-----FLSITVFPFLIPILEV-LNICVFPREVTNFLRKSVKRMKES---- 223 (445)
    Q d2j0da1
                                                          156 \quad \text{number of } \\ \text{number on } \\ \text{numb
    O Consensus
                                                                    165 \\ \text{ when } \hat{f} \\ \text{ for the production of the production of the production } \hat{f} \\ \text{ when } \hat{f} \\
                                                                                                                                                                                                                                                                                                                                244 (479)
    T Consensus
                                                           165 ASLTCGRRFEYDDPRFLRLLDLAQEGLKEESGFLREVLNAVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTWDPAQ 244 (479)
    T 3tbg A
                                                                       T ss_dssp
                                                                       T ss_pred
    Q ss_dssp
                                                                       Q ss_pred
    Q d2j0da1
                                                           224 -R--FLQLMIDSQNS-----HKALSDLELVAQSIIFIFAGYETTSSVLSFIMYELATHPDVQQKLQEEIDAVLPNKAPPT
                                                                                                                                                                                                                                                                                                                                295 (445)
                                                          224 \quad \text{$---1$} \\ \text{$-100$} \\ \text{$100$} \\ \text{$---100$} \\ \text{$100$} 
    O Consensus
                                                                                                                                                                                                                                                                                                                                295 (445)
                                                                       324 (479)
                                                           T Consensus
                                                           245 PPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGLLLMILHPDVQRRVQQEIDDVIGQVRRPE
    T 3tbg_A
                                                                       T ss_dssp
                                                                       T ss_pred
    Q ss_dssp
                                                                       HHHHTTCHHHHHHHHHHHHCCTTCC-EEEECCSCEECSSCEECTTCEEEECHHHHTTCTTTSSSTTSCCGGGGSTTTGG
                                                                       Q ss_pred
     Q d2j0da1
                                                          296 YDTVLQMEYLDMVVNETLRLFPIAMR-LERVCKKDVEINGMFIPKGVVVMIPSYALHRDPKYWTEPEKFLPERFSKKNKD
                                                                                                                                                                                                                                                                                                                             374 (445)
                                                          296 nnlnlnlaliaeinEnlRlnnnnn-nRnnndnnngnipngtnynnnnnndnnnnpnFnPnRnn
    Q Consensus
                                                                                                                                                                                                                                                                                                                               374 (445)
                                                                       .+----||||+||+||+||+||+---. -+-|.+.+||---||.||||||.||----.+---|||||+||+||+||+|||----.
                                                          T Consensus
                                                           325 MGDQAHMPYTTAVIHEVQRFGDIVPLGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQGH
    T 3tbg_A
                                                                                                                                                                                                                                                                                                                              404 (479)
    T ss_dssp
                                                                       HHHHTTCHHHHHHHHHHHHCSSTTCCCEECSSCEEETTEEECTTCEEEEEHHHHHTCTTTSSSTTSCCGGGGBCTTCC
    T ss_pred
                                                                       Q ss_dssp
                                                                       ccccccccccccchhнниннинниннинниннеевевсссс-ccceeeccccceesesesesec
    Q ss_pred
     Q d2j0da1
                                                          375 NIDPYTYTPFGSGPRNCIGMRFALMNMKLALTRVLQNFSFKPCKET-QIPLKLSLGGLLQPEKPVVLKVESR
                                                                                                                                                                                                                                                                                                       445 (445)
                                                          375 mannana Feaga Pa Cagana Annanana la la farananana - nana
                                                                                                                                                                                                                                                                                                       445 (445)
    Q Consensus
                                                                                                                                                                                                                                                     ~~~D~~~~V~~~~~
                                                                        ...+,+++|||,|+|,|+|++||+,|++++++,|++++,...+,...+,...+,....| ,++,|++++|
```

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.

Para correr modeller usamos el sigioente *script* de *python*:

 $https://github.com/carinapaola/Bioinformatica_estructural/blob/master/ModellerScript.py \\ Calidad para el modelo1:$

<<< end of ENERGY.

DOPE score :-141886.689193

Total CPU time [seconds]

Calidad para el modelo2:

<<< end of ENERGY.

DOPE score

Total CPU time [seconds]

:-129091.742344

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.

modelo1 vs d2j0da1

```
PSI(ini)= 94.95 NALI= 432 NORM= 455 RMS=
                                4.27 NSS= 377
PSI(end)= 86.81 NALI= 395 NORM= 455
                           RMS=
                                3.18
Sstr(LG)= 6178.25 NALI= 395 NORM= 455 RMS=
                                3.18
E-value=
        0.10299921E-17
Z-score= 43.806233
                -ln(E)= 41.416981
 Final Structural Alignment
           **** ******** *** *** ******* *****
Prediction .....GKLPP GPLPLPGLGN LLHVDFONTP YCFDOLRRRF GDVFSLOLAW
Experiment SSSSS---- -SSS----HH HHH-HHHHHH HHHHHHH-- ---SSSS---
Experiment HGLFKKLGIP GPTPLPFLGN ILS.YHKGFC MFDMECHKKY GKVWGFYDGQ
           -
**** ******* *** **** ***** ***
       ******* ****** * *******
Prediction TPVVVLNGLA AVREALVT.H GEDTADRPPV PITOILGFGP RSOGVFLARY
Prediction --SSSS---- HHHHHHHH-H HHHHHHHHHH HHH---SSS- -SSS-----
       Experiment -SSSSSS--- HHHHHHHHHH ---SSSSSS- ----S-SSSS--S-SS
Experiment OPVLAITDPD MIKTVLVKEC YSVFTNRRP. ....FGPVGF MKSAISI.AE
       ******* ****** * *******
       *******
                *** ** * ***** ****** **** ****
Prediction GPAWREQRRF ....SVSTLR NLGLGKKSLE QWVTEEAACL CAAFANHSGR
Experiment DEEWKRLRSL LSPTFTSGKL K...EMVPII AQYGDVLVRN LRRE.AETGK
                 *** ** * ***** ****** ****
       ******* ****** * *** ***
                                   ******
Prediction PFRPNGLLDK AVSNVIASLT C....GRRFE YDDPRFLRLL DLAOEGLKEE
Experiment PVTLKDVFGA YSMDVITSTS FGVNIDSLNN PQDPFVE... NTKKLLRFDF
       ******* ****** * *** ***
```

```
# total residuos: pdb1 = 445 pdb2 = 472
# total residuos alineados = 413
# coordenadas originales = original.pdb
# superposicion optima:
# archivo PDB = align_fit.pdb
# RMSD = 10.46 Angstrom
```

 ${\rm modelo2~vs~d2j0da1}$

```
PSI(ini)= 96.85 NALI= 431 NORM= 445 RMS=
                                4.67 NSS= 370
PSI(end)= 87.87 NALI= 391 NORM= 445
                           RMS=
                                3.35
Sstr(LG)= 5975.17 NALI= 391 NORM= 445 RMS=
                                3.35
E-value= 0.11926224E-17
Z-score= 43.646737 -ln(E)= 41.270377
 Final Structural Alignment
           **** ******* *** *** ****** *****
Prediction .....GKLPP GPLPLPGLGN LLHVDFQNTP YCFDQLRRRF GDVFSLQLAW
Experiment SSSSS---- -SSS----HH HHH-HHHHHH HHHHHHH-- ---SSSS---
Experiment HGLFKKLGIP GPTPLPFLGN ILS.YHKGFC MFDMECHKKY GKVWGFYDGQ
           -
************ ******** ****** ***
       ******* ****** * ******* * ****
Prediction TPVVVLNGLA AVREALVT.H GEDTADRPPV PITQILGFGP RSQGVFLARY
Prediction --SSSS---- HHHHHHHH-H HHHHHHHHHH HHH---SSS- -SSS-----
       П
Experiment -SSSSSS--- HHHHHHHHHH --SSSSSSSS ----SSSS- -S----SS
Experiment QPVLAITDPD MIKTVLVKEC YSVFTNRRPF GPVGFMKSAI SI.....AE
       ******* ****** * ******* * *******
                ** ***
                        * **** ******* ****
       *******
Prediction GPAWREORRF ....SVSTLR NLGLGKKSLE OWVTEEAACL CAAFANHSGR
Experiment DEEWKRLRSL LSPTFTSGKL K...EMVPII AQYGDVLVRN LRREA.ETGK
       *******
                 ** ***
                        * **** ******* ***** ****
       ******* ** ****** ****** **
Prediction PFRPNGLLDK AVSNVIASLT CGRRFEYDDP RFLRLLDLAQ EGLKEESGFL
Experiment PVTLKDVFGA YSMDVITSTS FGVNIDSNPQ DPFVENTKK. ....LLRFF
       ******** ******* ****** ***
```

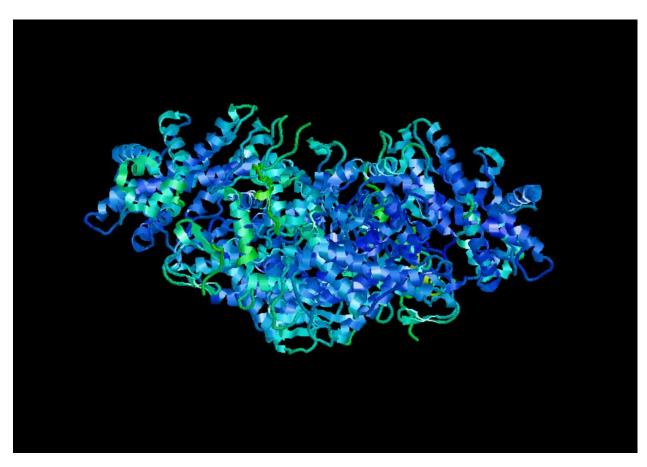
```
# total residuos: pdb1 = 445 pdb2 = 472
# total residuos alineados = 413

# coordenadas originales = original.pdb
# superposicion optima:
# archivo PDB = align_fit.pdb
# RMSD = 8.76 Angstrom
```

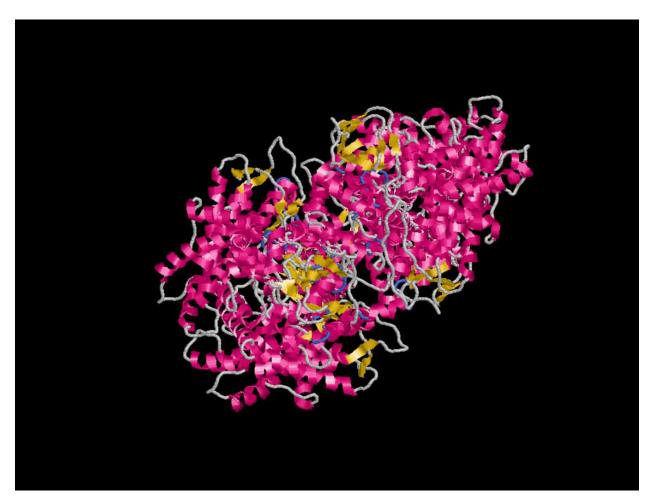
d2j0da1



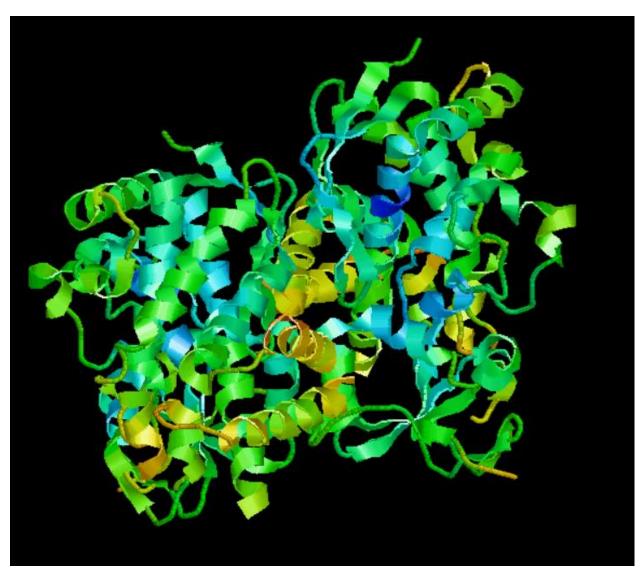
modelo1



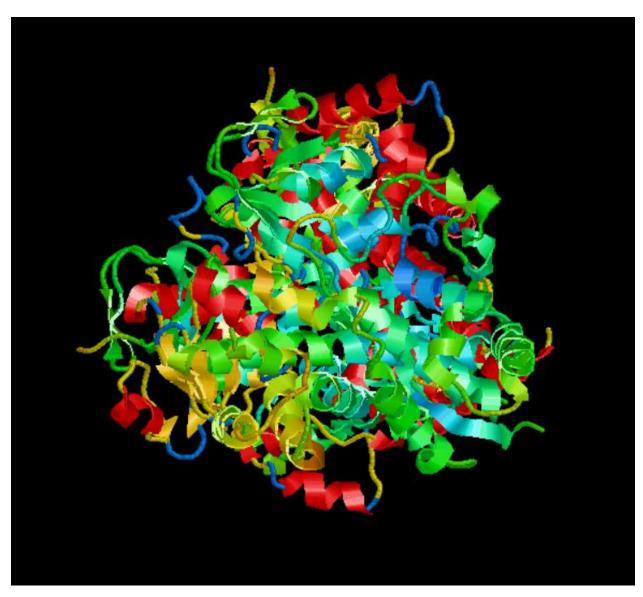
d2j0da1 vs modelo1



modelo2



d2j0da1 vs modelo2



Se aprecia que los modelos si son diferentes del citocromo P450 d2j0da1, suponemos que se debe a que nuestro templete tiene solo 22% de identidad con la proteína query, escojimos dicho templete porque el resto de los posibles templetes tenian valores muy cercanos, todos alrededor del 20% de identidad, o tenian el 100% de identidad, pudiendo ser la misma proteína. A simple vista los modelos no parecen exactamente iguales a la proteína d2j0da1, sin embargo los valores DOPE, que nos indican la calidad del modelo es negativa, lo cual indica que son buenos modelos.

El DOPE score del modelo1 es más negativo, además su RMSD es ligeramente mayor que la del modelo2, por lo que asumimos que el modelo 1 es un mejor modelo de nuestra proteína.