

Practica3

Carmina Barberena Jonas

1. Selecciona una superfamilia de proteínas de SCOP (<http://scop.berkeley.edu>) y extrae la secuencia de aminoácidos (ATOM records) y las coordenadas PDB de varios dominios de la misma. Podéis ver un ejemplo de dominio en <http://scop.berkeley.edu/sunid=29763>, y abajo están tanto la secuencia como una liga para descargar las coordenadas.

Class a: All alpha proteins.

Superfamily a.4.1: Homeodomain-like.

-Species *Saccharomyces cerevisiae*

-Species *Escherichia coli*

-Species Fission yeast (*Schizosaccharomyces pombe*)

-Species *Bacillus subtilis*

Secuencia de aa (ATOM records) >d2fq3a1 a.4.1.18 (A:311-395)
Transcription regulatory protein swi3 {Baker's yeast (*Saccharomyces cerevisiae*) [TaxId: 4932]}
skwfnlekihsievqslpefftnripsktpevymryrnfmvnsyrlnpneyfsvtarrn
vsgdaaalfrlhkfltkwglinyqv

d1gdta1 a.4.1.2 (A:141-183) gamma,delta resolvase (C-terminal domain)
{*Escherichia coli* [TaxId: 562]} grkrkidrdavlnmwqqglgashisktmniarstvykvinesn

d1iufa2 a.4.1.7 (A:76-141) Ars-binding protein 1, ABP1 {Fission yeast (*Schizosaccharomyces pombe*) [TaxId: 4896]}
ppkypllaalfewvqqgddatlsgetikraaailwhkipeyqdpvpnfngwlegfr krhilh

d2o3fa1 a.4.1.20 (A:1-83) Putative transcriptional regulator YbbH {*Bacillus subtilis* [TaxId: 1423]}
matgglaaiqsmkhklppserkladyilahphkaiestvneisalanssdaavirlcksl
glkgfqlkmrvagdlakptfqq 2) Comprueba que las secuencias descargadas coinciden con las coordenadas. ##Script para comprobar que esten correctamente anotados

```
#!/usr/bin/perl -w
use strict;
my @AA; ##Variablepara almacenar aa en codigo 3 letra
my @AA1; ##Variablepara almacenar aa en codigo letra
my %HAA1=('CYS'=>'C', 'ASP'=>'D', 'SER'=>'S', 'GLN'=>'Q', 'LYS'=>'K',
'ILE'=>'I', 'PRO'=>'P', 'THR'=>'T', 'PHE'=>'F', 'ASN'=>'N',
'GLY'=>'G', 'HIS'=>'H', 'LEU'=>'L', 'ARG'=>'R', 'TRP'=>'W',
'ALA'=>'A', 'VAL'=>'V', 'GLU'=>'E', 'TYR'=>'Y', 'MET'=>'M'); #hash con
```

```

los codigos de aa
my $infile1 = $ARGV[0] || die "# usage: $0 <promoters file>\n";
my $infile2 = $ARGV[1] || die "# usage: $0 <promoters file>\n";
open(SEQ1, $infile1) || die "#cannot open input $infile1 : $!\n";
while(<SEQ1>) #archivo 1 a parsear
{
    if(/^ATOM.{9}CA.{2}(\w+)/){ #Expresión regular para encontrar el carbono
    alfa
        push @AA, $1; #Almacenarlo en un hash
    }
}
close(SEQ1);
for (my $n=0; $n<(scalar(@AA));$n++){
    push @AA1 , $HAA1{@AA[$n]}; #Convertir en codigo 1 letra
}
#print scalar(@AA)."otro".scalar(@AA1); #Comprobar
print join("", @AA1)."\n";
open(SEQ2, $infile2) || die "#cannot open input $infile2 : $!\n";
while(<SEQ2>) #Archivo a comparar
{
    if (/^(\w+)/){
        print "$1";
    }
}
close(SEQ2);

```

Con esto podemos corregir las secuencias manualmente.

Resultados: Las primeras 3 secuencias resultaron estar correctas pero la ultima mostro no tener concordancia por lo cual se va a eliminar

```

cbarbere@cbarbere-Lenovo-G575:~/Bioinfo/Bruno/Pratica 2$ perl ParseoAA.pl d2o3fa1.pbb d2o3fa1.fa
ATGGLAIQSHLPPSEERKLADYILAHPHAIESTVNEISALANSSDAVIRRLCSLGLKGFQDLRVAGDLAKPTFQG
matgglaiqsmkhklppserkladyilahphkaiestvneisalanssdaavirllcksllglkgfqlkrmrvagdlakptfqgcbarbere@cbarbere
Bioinfo/Bruno/Pratica 2$

```

- 3 Calcula al menos dos alineamiento pareados entre secuencias de aminoácidos de las extraídas en 1 y calcula su %identidad como el total de parejas de residuos idénticas / total parejas alineadas. Compararemos la secuencia de Pombe contra Coli y Sacaro

Pombe contra Coli

Identidad =42%



d1iufa2 a.4.1.7 (A:76-141) Ars-binding protein 1, ABP1 {Fission yeast (Schizosaccharomyces pombe) [TaxId: 4896]}
Sequence ID: Icl|Query_236687 Length: 66 Number of Matches: 1

Range 1: 53 to 64 Graphics			▼ Next Match ▲ Previous Match		
Score	Expect	Method	Identities	Positives	Gaps
14.2 bits(25)	0.063	Compositional matrix adjust.	5/12(42%)	6/12(50%)	0/12(0%)
Query	13	NMWQQGLGASHI	24		
		N W +G HI			
Sbjct	53	NGWLEGFRKRHI	64		

Pombe contra Sacaromi

Identidad=18%

d2fq3a1 a.4.1.18 (A:311-395) Transcription regulatory protein swi3 {Baker's yeast (Saccharomyces cerevisiae) [TaxId: 4932]}
Sequence ID: lcl|Query_46287 Length: 85 Number of Matches: 1

Range 1: 6 to 27 Graphics				 Next Match  Previous Match		
Score	Expect	Method		Identities	Positives	Gaps
13.5 bits(23)	0.32	Compositional matrix adjust.		4/22(18%)	12/22(54%)	0/22(0%)
Query 29	IKRAAAILWHKIPEYQDQVPVN 50					
	+++ +I +PE+ +P+					
Sbjct 6	LEKIHISIEVQSLPEFFTNRIPS 27					

- 4 Calcula con mammoth los alineamientos estructurales de los dominios que ya alineaste en 3 en base a su secuencia. Visualízalos con Rasmol como se explica en http://eead-csic-compbio.github.io/bioinformatica_estructural/node32.html. El software está en /home/compu2/algoritmos3D/soft/mammoth-1.0-src para que lo copien y compilen con gfortran como se explica en README, cambiando g77 por gfortran.

Comandos

Pombe vs Sacaromisse

```
-bash-3.00$ mammoth -p dliufa1.pdb -e d2fq3a1.pdb -o log.out
```

Pombe vs E.Coli

```
-bash-3.00$ mammoth -p dliufa1.pdb -e d1gdta1.pdb -o log.out
```

Resultados

Se generaron los siguientes archivos

```
-rw----- 1 cbarbere students 69951716 Mar  6 21:43 core
-rw-r--r-- 1 cbarbere students  28203 Mar  6 21:55 d1gdta1.pdb
-rw-r--r-- 1 cbarbere students 108810 Mar  6 21:35 dliufa1.pdb
-rw-r--r-- 1 cbarbere students  58904 Mar  6 21:35 d2fq3a1.pdb
-rw-r--r-- 1 cbarbere students  52664 Mar  6 21:35 d2o3fa1.pdb
-rw-r--r-- 1 cbarbere students  2100 Mar  6 21:55 log.out_Pombe_vs_Ecoli
-rw-r--r-- 1 cbarbere students  2191 Mar  6 21:44 log.out_Pombe_vs_Saca
-rw-r--r-- 1 cbarbere students  7046 Mar  6 21:55
maxsub_sup.pdb_Pombe_vs_Coli
-rw-r--r-- 1 cbarbere students  9356 Mar  6 21:44
maxsub_sup.pdb_Pombe_vs_Saca
-rw-r--r-- 1 cbarbere students  6553 Mar  6 21:55
maxsub_sup2.pdb_Pombe_vs_Coli
-rw-r--r-- 1 cbarbere students  8863 Mar  6 21:44
maxsub_sup2.pdb_Pombe_vs_Saca
-rw-r--r-- 1 cbarbere students  7454 Mar  6 21:55 rasmol.tcl.Pombe_vs_Coli
-rw-r--r-- 1 cbarbere students  9506 Mar  6 21:44 rasmol.tcl.Pombe_vs_Saca
```

Los resultados que se encuentran en log.out ####Pombe vs Coli

Input information

==> PREDICTION:

Filename: d1iufa1.pdb
Number of residues: 78

==> EXPERIMENT:

Filename: d1gdta1.pbb
Number of residues: 43

Structural Alignment Scores

Filename: d1gdta1.pbb
Number of residues: 43

Structural Alignment Scores

PSI(ini)= 97.67 NALI= 42 NORM= 43 RMS= 5.90 NSS= 27
PSI(end)= 72.09 NALI= 31 NORM= 43 RMS= 3.98
Sstr(LG)= 447.42 NALI= 31 NORM= 43 RMS= 3.98

E-value= 0.20191117E-01

Z-score= 3.7071971 -ln(E)= 3.9025125

Final Structural Alignment

***** * * * ***** * *
Prediction GIHMGKIKRR AITEHEKRAL RHYFFQLQNR SGQQDLIEWF REKFGKDISQ
Prediction SSSSSS--- HHHHHHHHHHH HHHHHH---- HHHHHHHHHHH HH---
HHHHH
| ||||| ||| ||||| |||
Experiment -----HHH HHHHHHHHHHH -----HH--- H----HHHHH HH-----HHH

ExperimentGRK RKIDRDAVLNMWQQG L....GASHI SKT..MNIAR
***** * * * ***** * * *

Prediction PSVSQILSSK YSYLDNTVEK PWDVKRN
Prediction HHHHHHHHHH HHHHHHHHHH HHHHHHHH
|||
Experiment HHHHHHHHHH -----
Experiment STVYKVINES

Pombe vs Sacaro

Input information

==> PREDICTION:

Filename: d1iufa1.pdb
Number of residues: 78

==> EXPERIMENT:

Filename: d2fq3a1.pdb
Number of residues: 85

Structural Alignment Scores

PSI(ini)= 92.31 NALI= 72 NORM= 78 RMS= 12.60 NSS= 40
PSI(end)= 32.05 NALI= 25 NORM= 78 RMS= 3.04
Sstr(LG)= 458.78 NALI= 25 NORM= 78 RMS= 3.04

E-value= 0.15917422

Z-score= 1.4215475 -ln(E)= 1.8377560

Final Structural Alignment

***** ***** *** *****
Prediction GIHMGKIKRR AITEHEKRAL RHYFFQLQNR SGQQDLIEWFREK
Prediction SSSSSS--- HHHHHHHHHH HHHHHH--- HHHHHHHHHH -----HH-

```

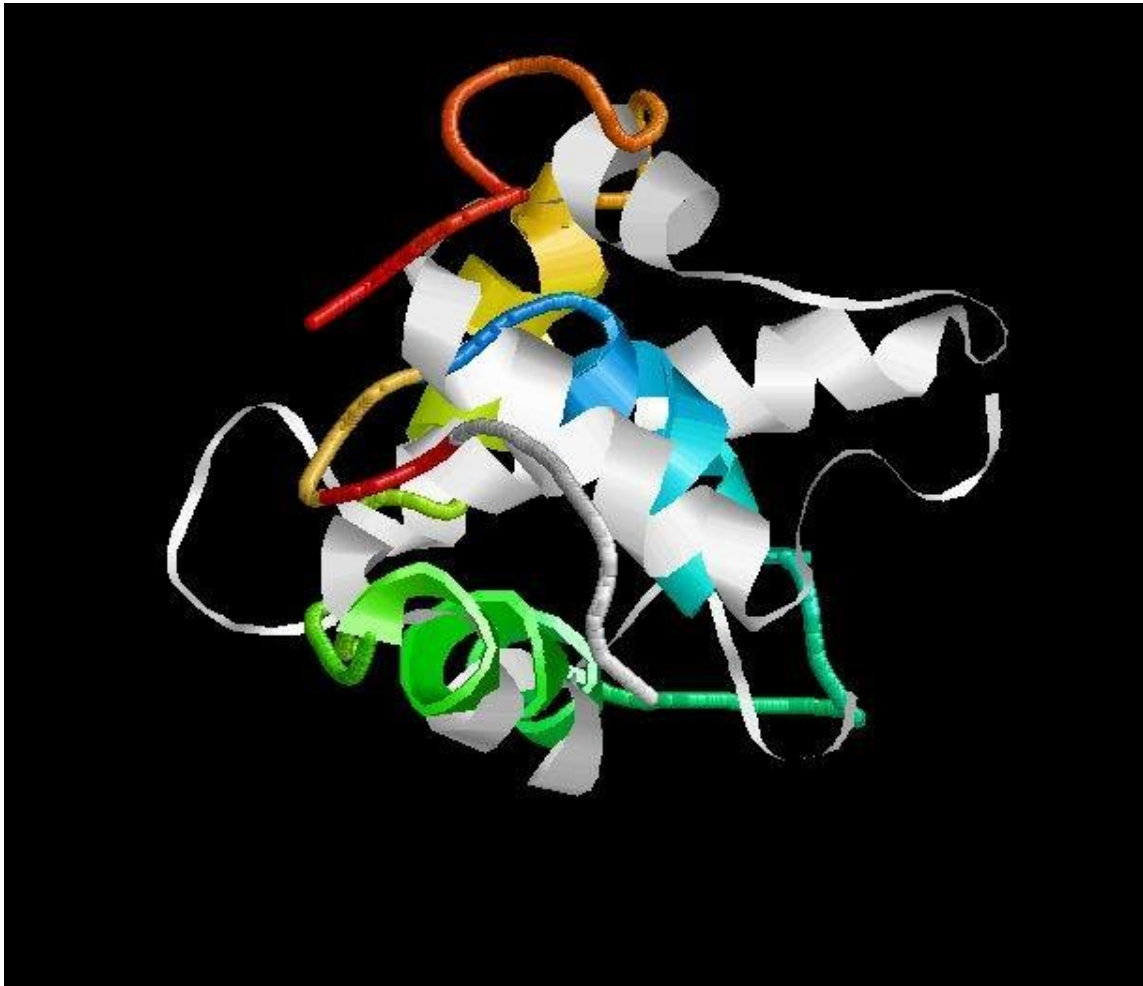
      ||||| |||||      | ||||| |||      |
Experiment ---SSSSS-- -HHHHHHHHHH HHHHH----- HHHHHHHHHHH
HHHHHHHHHH-
Experiment ...SKWFNLE KHSIEVQSL PEFFTNRIPS KTPEVYMRYR NFMVNSYRLN
      ***** ***** ***          *****

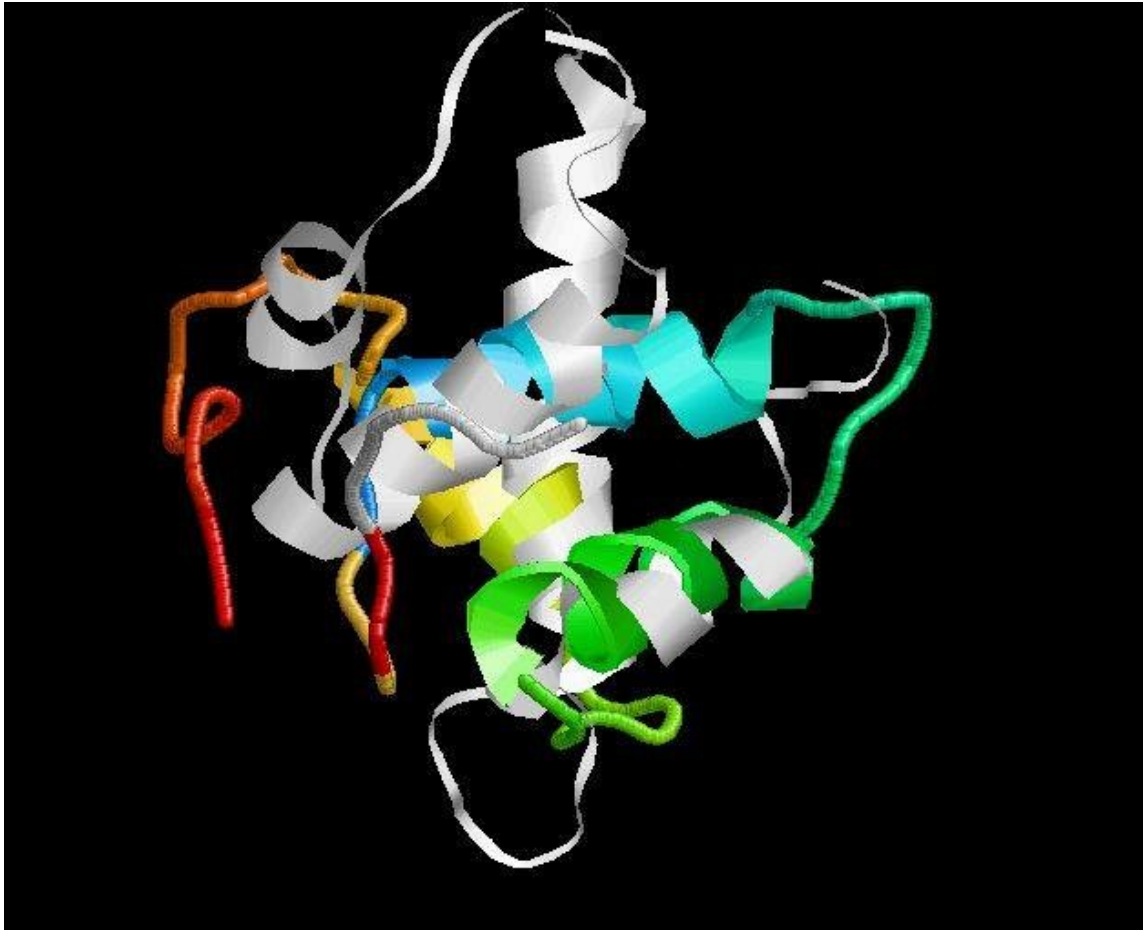
Prediction FGKDISQPS. ....VSQILS SKYSYLDNTV EKPWDVKRN
Prediction --HHHHHHH- ----HHHHHH HHHHHHHHHHH HHHHHHHHHH
      ||||| ||| |||      |
Experiment ---HHHHHHH HH---SS--H HHHHHHHHHHH -SSSSS--
Experiment PNEYFSVTTA RRNVSGDAAA LFRLHKFLTK WGLINYQ..

```

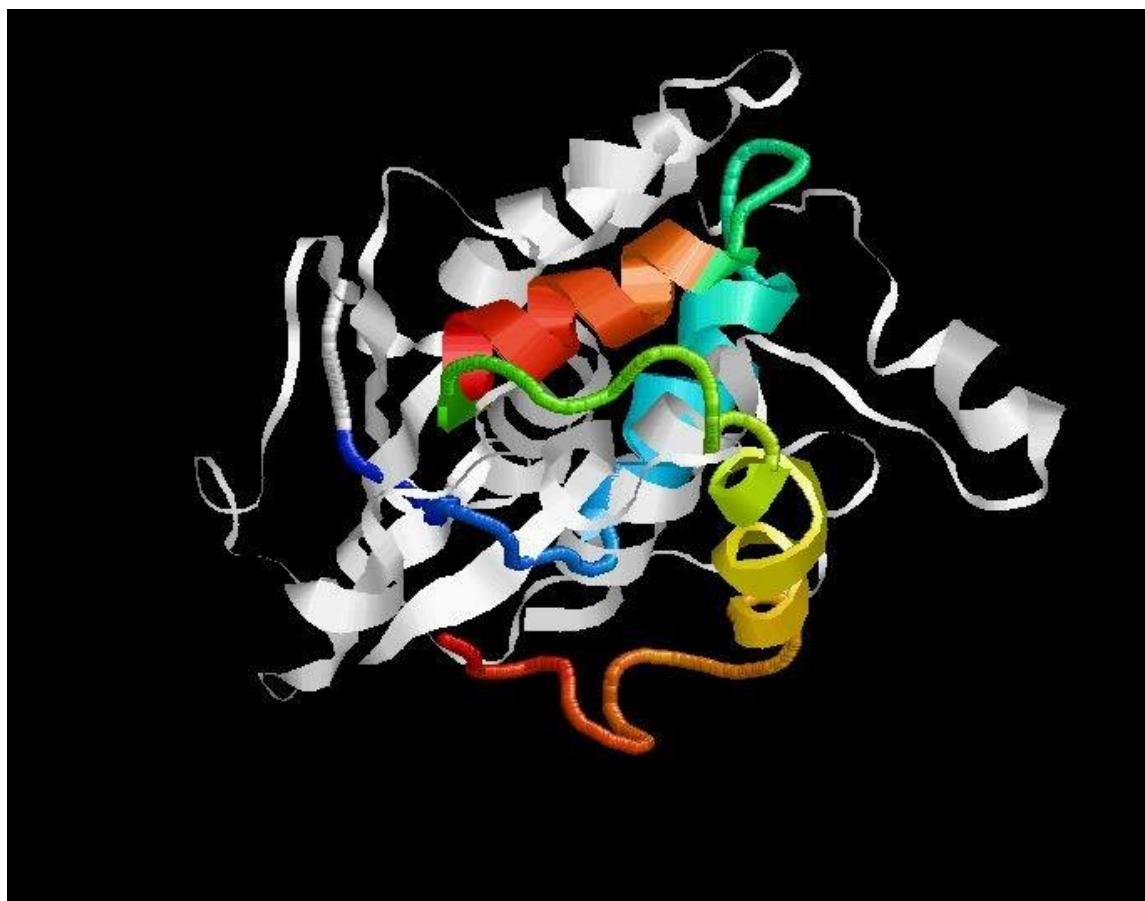
Rasmol

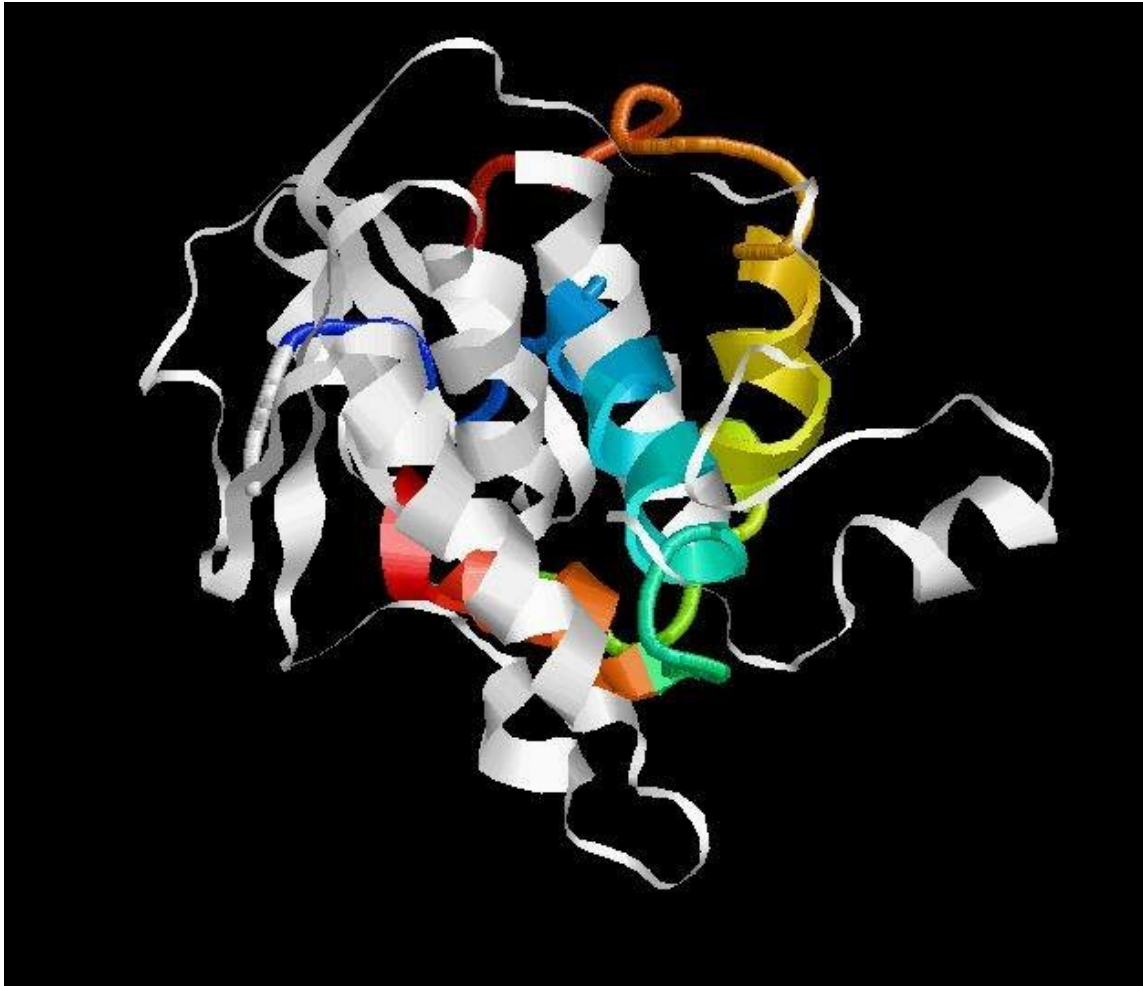
Pombe vs Sacara





Pombe vs E.coli





- 5 Compara los alineamientos obtenidos en 3 y 4. Comenta en qué elementos de estructura secundaria se observan diferencias.

Con respecto a la estructura secundaria se pueden apreciar varias helices lo cual concuerda ya que estamos en la super familia que contiene unicamente helices. Aunque varios aminoacidos no se alinean en secuencias (3) podemos ver como en estructura secundaria (4) contienen la misma conformación. Esto es muy interesante ya que aunque la identidad no es tan grande siguen compartiendo un rasgo importante que es la estructura y muy posiblemente la función. Esto se observa en mayor medida en la comparación de Pombe y Sacaromicces. Al visualizarlos en 3d podemos ver que algunos alfa helice se acoplan mientras que otros varían pero en general se mantiene la estructura.

- 6 Utiliza el prog3.1 (en http://eead-csic-compbio.github.io/bioinformatica_estructural/node31.html) para calcular el error (RMSD) de los alineamientos obtenidos en 3 y 4 y comenta los resultados. Son mejores o peores los alineamientos basados en secuencia desde el punto de vista del RMSD?

Aqui se muestran los resultados del punto 3

```
print "\n# coordenadas originales = original.pdb\n# superposicion optima:\n",
print "# archivo PDB = align_fit.pdb\n# RMSD = %1.2f Angstrom\n" % (rmsd),

# total residuos: pdb1 = 78 pdb2 = 85
# total residuos alineados = 40

# coordenadas originales = original.pdb
# superposicion optima:
# archivo PDB = align_fit.pdb
# RMSD = 5.65 Angstrom

In [12]: runfile('C:/Users/Jess/Documents/Python Scripts/PROYECTO BRUNO.p
wdir='C:/Users/Jess/Documents/Python Scripts')
Reloaded modules: SVD
# total residuos: pdb1 = 78 pdb2 = 43
# total residuos alineados = 43
|
# coordenadas originales = original.pdb
# superposicion optima:
# archivo PDB = align_fit.pdb
# RMSD = 5.79 Angstrom
```

Los alineamientos en el punto 4 obtienen mejores puntajes de RMSD esto debido a que estos estan tomando la conformacion secundaria, lo cual nos proporciona una mayor cantidad de informacion ya que como se menciono anteriormente se pueden tener datos mas importantes con respecto a la proteinas, ya que pueden modificar su secuencia pero mantener su estructura que es lo que las hace funcionales.