

Bioinformática II
Algoritmos 3D-3
Integrantes:
Cruz Ruiz Jessica Samantha
Fajardo Brigido Lorena Elizabeth

Las acuaporinas son proteínas fundamentales de la membrana en las células de los organismos, estas son importantes ya que controla los flujos de agua a través de la membrana y su osmolaridad interna. Son importantes ya que se ha descubierto que cuando se tienen acuaporinas defectuosas estas conducen a enfermedades, como cataratas, edemas pulmonares, glaucoma, entre otros.

1) Selecciona una superfamilia de proteínas de SCOP

La superfamilia de proteínas que seleccionamos fue la de Aquaporin-like, de la cual elegimos 3 proteínas.

Proteína	Organismo	Dominio
Aquaporina Z	E. coli	d1rc2a_
Aquaporina 0	Bos taurus	d1ymga1
Aquaporina 1	Homo sapiens	d1ih5a_

2) Comprueba que las secuencias descargadas coinciden con las coordenadas

Para cada una de estas proteínas obtuvimos los formatos ATOM y PDB, de la página SCOP. Para corroborar que cada una de las secuencias y sus coordenadas estuvieran bien anotadas y fueran las mismas del archivo formato ATOM obtuvimos la secuencia de la proteína para compararla con la secuencia del archivo fasta.

Todas las secuencias y sus coordenadas coincidieron, por lo cual procedimos con los siguientes pasos.

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.

Realizamos alineamientos pareados entre las secuencias de aminoácidos con la herramienta BLAST, donde los alineamientos se realizaron de la siguiente forma;

Organismos	Porcentaje de Identidad
E. coli - Homo sapiens	77/213(36%)
Bos Taurus - E. coli	81/230(35%)
Homo sapiens - Bos Taurus	108/222(49%)

Se muestran a continuación los alineamientos de BLAST.

E. coli - Homo sapiens

D1RC2A_F.19.1.1 (A:) AQUAPORIN Z {ESCHERICHIA COLI [TAXID: 562]}
Sequence ID: Icl|Query_28421 Length: 231 Number of Matches: 1

Range 1: 3 to 209 [Graphics](#)

Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
98.2 bits(243)	4e-29	Compositional matrix adjust.	77/213(36%)	109/213(51%)	17/213(7%)
Query 4	RAVVAEFLATTLFVFISIGSALGFKYPVGNQTAVQDNVKSALAFGLSIATLAQSVGHIS				63
Sbjct 3	R + AE T VF GSA+ G + + V+LAFGL++ T+A +VGHIS				58
Query 64	GAHLNPAVTLGLLSCQISIFRALMYIIAQCVGAIVATAILSGITSSLTGNLSGRNDLAD				123
Sbjct 59	G H NPAVT+GL + + Y+IAQ VG IVA A+L I S TG + A				118
Query 124	---GVNSGQG-----LGIEIIGTLQLVLCVLATDRRRDLGGSAPLAIGLSVALGHLL				174
Sbjct 119	G +S G L +E++ + +L + TD+ G AP+ATGL++ L HL+				176
Query 175	AIDYTGCGINPARSFGSAVITHNFS--NHWIFW 205				
Sbjct 177	+I T +NPARS A+ ++ W FW				209
	SIPVTNTSVNPARSTAVAIFOGGWALEQLWFFW				

Bos Taurus - E. coli

D1YMGAI_F.19.1.1 (A:6-239) AQUAPORIN-0 {COW (BOS TAURUS) [TAXID: 9913]}
Sequence ID: Icl|Query_20987 Length: 233 Number of Matches: 1

Range 1: 6 to 220 [Graphics](#)

Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
105 bits(263)	7e-32	Compositional matrix adjust.	81/230(35%)	116/230(50%)	17/230(7%)
Query 3	RKLAAECFGTFLVFGGCGSAVLAAGFPPELGIFAGVALAFGLTVLTMFAVGHISGGHF				62
Sbjct 6	R + AE F + + VF G G+++ AG + VALAFGL + T+ AVGHISG H				61
Query 63	NPAVTIGLWAGRRFPAKEVVGVIQVVGIVAAALLYIASGKTGFDAASGFASNGYG				122
Sbjct 62	NPAVTG + + + Y++AQ++G + AA+LY + A A N				116
Query 123	EHSPGGYSMLSALVELVLSAGFLVTHGATD--KFAPAGFAPITGLALTLIHLISIPV				180
Sbjct 117	P----GVSVGQATIVEIFLTQLQVLCIFATYDERRNRLGSVALAVGFSLTGLHFGMY				172
Query 181	TNTSVNPARSTAVAIFOGNALEQLWFFWVPIVGGIIGGLYRTLEKR 230				
Sbjct 173	T +NPARS A AI + W +WV P++G +G L+Y LL R				220
	TGAGMNPARSFAPAILTRNF--TNHWYVWGPVIGAGLGSLLYDFLLFPR				

Homo sapiens - Bos Taurus

D1YMGAI_F.19.1.1 (A:6-239) AQUAPORIN-0 {COW (BOS TAURUS) [TAXID: 9913]}
Sequence ID: Icl|Query_137413 Length: 233 Number of Matches: 1

Range 1: 4 to 217 [Graphics](#)

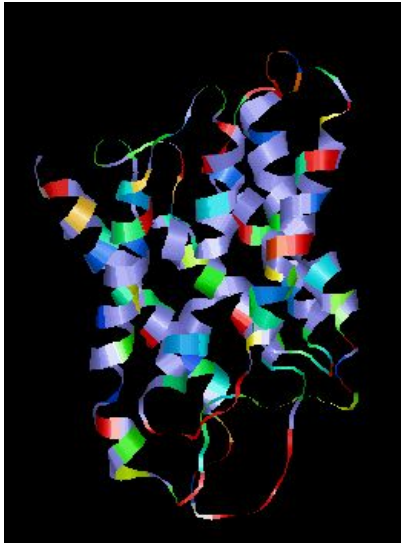
Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
220 bits(560)	6e-76	Compositional matrix adjust.	108/222(49%)	148/222(66%)	8/222(3%)
Query 2	FWRRAVVAEFLATTLFVFISIGSALGFKYPVGNQTAVQDNVKSALAFGLSIATLAQSVGH				61
Sbjct 4	FWRRA+ AEF A+ +VF +G++L + P+ ++V+LAFGL++ATL Q+VGH				55
Query 62	ISGAHLNPAVTLGLLSCQISIFRALMYIIAQCVGAIVATAILSGITSSLTGNLSGRNDL				121
Sbjct 56	ISGAH+NPAVT L+ Q+S+ RA+ Y++AQ +GA+ A+L +T +L N L				115
Query 122	ADGVNSGQGLGIEIIGTLQLVLCVLATDRRRDLGGSAPLAIGLSVALGHLLAIDYTG				181
Sbjct 116	HPGVSVGQATIVEIFLTQLQVLCIFATYDERRNRLGSVALAVGFSLTGLHFGMYTGA				175
Query 182	GINPARSFGSAVITHNFSNHWIFWVGPFGGALAVLIYDFIL 223				
Sbjct 176	G+NPARSF A++T NF+NHW++WVGP IG L L+YDF+L				217
	GMNPARSFAPAILTRNFTNHWVWVGPVIGAGLGSLLYDFLL				

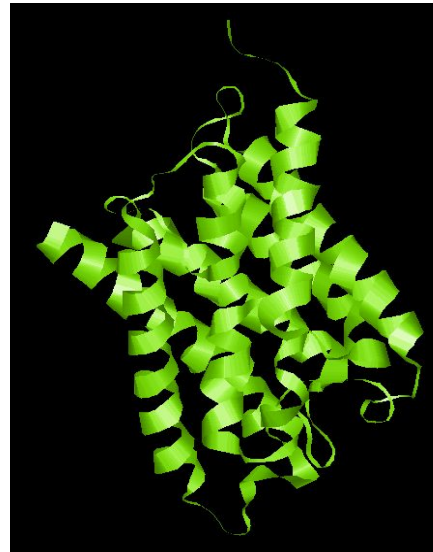
Se visualizaron las proteínas con el programa RasMol, en el cual pudimos interactuar con la estructura terciaria de las proteínas y notar que las estructuras entre proteínas de la misma familia son muy parecidas entre sí.

4) Calcula con mammoth los alineamientos estructurales de los dominios que ya alineaste en 3 en base a su secuencia.

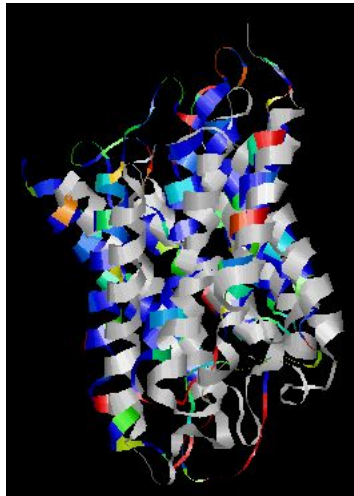
E. coli - Homo sapiens



D1IH5A_ (Homo sapiens)

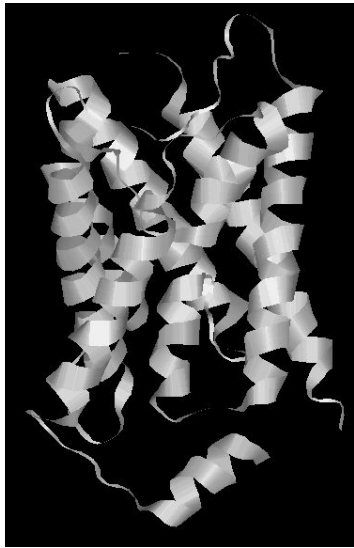


D1RC2A_ (E. coli)

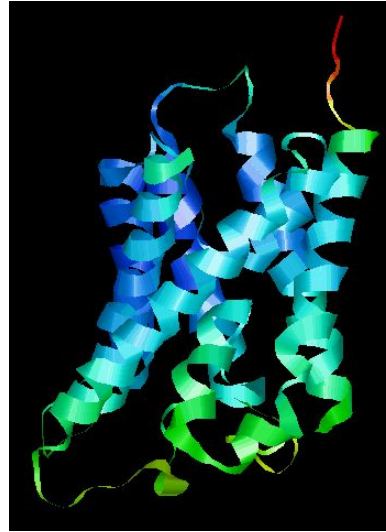


En estas dos proteínas podemos observar que las estructuras no son muy parecidas, pero entre ellas existe el mismo tipo de plegamientos. En la proteína de Homo Sapiens podemos ver que se nota una proteína con una estructura mas abierta, sin embargo entre ésta y la estructura de la proteína de E. coli se puede decir que es conservada.

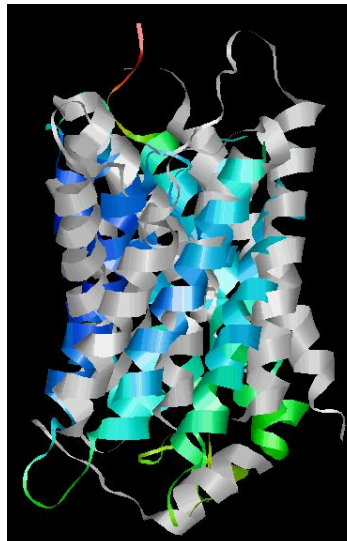
Bos Taurus - E. coli



D1MYGA **Bos Taurus**

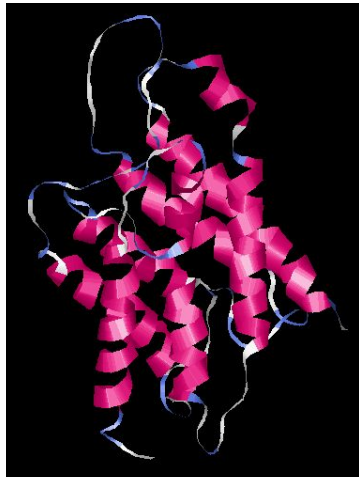


D1RC2A_ (E. Coli)

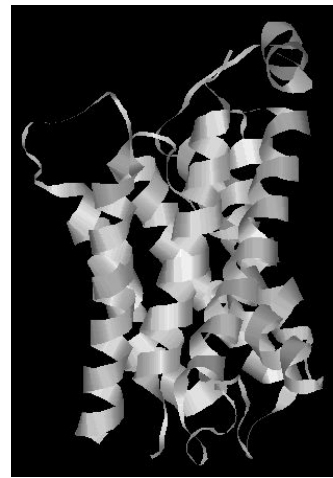


Entre estas dos estructuras de proteínas, se puede observar la misma densidad en la estructura, a comparación de las estructuras de E. coli - Homo sapiens, hay un patrón parecido entre los plegamientos de las dos estructuras, las estructuras son muy parecidas pero a pesar de eso sus plegamientos entre las alfa hélices y su posición dentro de la proteína es diferente.

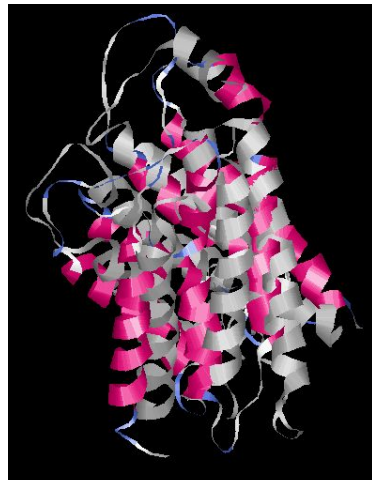
Homo sapiens - Bos Taurus



D1IH5A_ (Homo sapiens)



D1MYGA (Bos Taurus)



Estas estructuras de proteínas, tienen plegamientos más parecidos entre ellas que las dos anteriores, las estructuras se parecen por los grados de abertura que tienen entre las alfa-hélices y con los mismos plegamientos. Al ser organismos más cercanos se puede observar más semejanza, ya que tienen una distribución muy parecida.

En las estructuras de las tres proteínas se nota una gran similitud, incluso en organismos tan alejados como la bacteria *E. coli* y el humano, esto nos habla de que son proteínas conservadas entre reinos y por ende podemos deducir que muy importantes en los organismos.

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

Se realizaron dos alineamientos, como anteriormente se mencionó en BLAST y el otro en Mammoth, para cada una de los pares de las secuencias; por ejemplo de el par "*E. coli* - Homo sapiens", los E-value son diferentes en cada uno de los alineamientos, mientras en BLAST obtienen un E-value de $4E-29$, en mammoth obtienen un E-value $0.30841409E-10$. Los dos E-value obtenidos por cada uno de los programas es muy pequeño, lo que nos da un buen margen de confiabilidad.

6) Utiliza el prog 3.1 para calcular el error (RMSD) de los alineamientos obtenidos en 3 y 4 y comenta los resultados.

Realizamos la prueba de RMSD para certificar que tan buena eran los alineamientos a nivel de secuencia, para cada uno de los alineamientos;

Organismos	RMSD Angstroms
Homo sapiens - Bos Taurus	5.02 Angstrom
E. coli - Bos Taurus	5.72 Angstrom
E. coli - Homo Sapiens	5.76 Angstrom

Los valores obtenidos en cada uno de los alineamientos es muy bajo, lo que nos puede hablar de que los alineamientos a nivel de secuencia no son del todo certeros, ya que frecuentemente las secuencias pueden no ser homólogas, sin embargo el plegamiento del producto de esa secuencias es muy similar, y viceversa, puede darse el caso que la secuencia sea muy similar, sin embargo, la estructura de la proteína sea completamente diferente. Desde el punto de vista de los valores obtenidos por RMSD, los alineamiento a nivel de secuencia no son tan buenos.