

# Introducció

---

Em dic Laia Herrero Nogareda, visc a la Vall de Bianya, a la Garrotxa, i aquest curs 2010-2011 estic a l'Institut Montsacopa d'Olot cursant 2<sup>on</sup> de Batxillerat.

El treball de recerca que he realitzat està relacionat amb l'evolució de les proteïnes. Aquest és un camp d'estudi bàsic per entendre la biodiversitat del nostre planeta. A més, ara vivim en una bona època per avançar en el coneixement d'aquesta disciplina, ja que les tècniques de Biologia molecular han augmentat molt darrerament en nombre i precisió, i la Bioinformàtica ens permet veure les proteïnes ben de prop amb un detall impressionant.

Les proteïnes són molt importants en qualsevol organisme, ja sigui en forma d'hormones com d'enzims perquè permeten la realització de bona part de les funcions d'aquest organisme. I, com que són les majors responsables de les diverses funcions d'un organisme, una o diverses mutacions a l'ADN poden donar lloc a nous gens que codifiquin noves proteïnes amb noves funcions, fet que ens permet obtenir biodiversitat. Per tant, entendre com les proteïnes evolucionen és entendre les bases de l'evolució i l'especiació (aparició de noves espècies).

És per aquest motiu que el meu treball comença amb una part teòrica on explico com les mutacions porten l'evolució i l'especiació. Per la redacció d'aquesta part, vaig utilitzar un article bibliogràfic en anglès que jo havia realitzat a finals del 2009 pel programa de continuïtat de la beca d'Estades d'Estiu de Ciència de Caixa Catalunya, del programa Joves i Ciència de Caixa Catalunya. Gràcies a aquesta beca, vaig realitzar una estada a les Planes de Son, a la província de Lleida, de l'1 al 12 de juliol del 2009, duent a terme experiments en el projecte de Genètica i Biologia molecular.

La beca consta d'un programa de continuïtat i, per a poder-hi optar, havia d'escriure un article científic obligatòriament. Jo vaig redactar un article bibliogràfic titulat *How mutations drive evolution and speciation* amb la supervisió dels meus professors de projecte de les Planes de Son. Per tant, tenia ganes de poder realitzar una part pràctica del meu article per poder veure amb els meus propis ulls les conseqüències de les mutacions. D'aquesta manera, per a la part teòrica només vaig haver de traduir l'article al català, eliminar certs apartats i afegir-ne d'altres.

En el programa de continuïtat de la beca vaig participar en un camp internacional de ciència que organitza l'XLAB, un laboratori per a joves de la Universitat de Göttingen, a Alemanya. Allà vaig realitzar-hi una estada del 20 de juny al 13 de juliol del 2010, la qual es dividia en tres setmanes i, durant la primera, vaig dur a terme un projecte anomenat Biologia Molecular. En aquest projecte vam tractar els mecanismes de l'evolució de ben a prop, ja que vam utilitzar les mutacions i la selecció natural per aconseguir evolució. D'aquesta manera, seguint un experiment ja dissenyat, podia comprovar si el que havia explicat en la part teòrica era demostrable amb només quatre dies, que era la durada del projecte. I aquest experiment ocupa la segona part del meu treball.

El següent que em vaig proposar era agafar proteïnes que ja haguessin sofert un procés d'evolució i estudiar-les. Per a la realització d'aquest tercer apartat vaig comptar amb el suport del Parc Científic de Barcelona, ja que vaig ser acceptada en el seu Programa de "Recerca a Secundària", gràcies al qual comptava amb el suport de les seves instal·lacions i,

sobretot, amb el suport d'un tutor, que en el meu cas va ser l'Oliver Carrillo, físic, el qual em va ajudar a utilitzar les eines que necessitava pel meu treball i m'ajudava en qualsevol dubte.

Vaig acordar amb l'Oliver que podríem agafar la mateixa proteïna en diferents espècies, per tal d'estudiar-les i comparar-les utilitzant eines bioinformàtiques. Vam triar l'hemoglobina a causa de la seva gran popularitat, i les espècies estudiades van ser l'*Homo sapiens* (humà), la *Canis lupus familiaris* (gos), la *Anas platyrhynchos* (ànec de collverd), la *Columba livia* (colom) i la *Thunnus thynnus* (tonyina). A diferència de les pràctiques a Alemanya, les quals s'havien de realitzar en un laboratori ben equipat, aquestes les podia realitzar des de casa, només tenint un ordinador i els programes adequats. Per aquest motiu i pel fet que no estava limitada per una estada de pocs dies, aquesta part del treball va ser molt més creativa i divertida per a mi, ja que em vaig poder expandir tant com vaig voler i vaig poder estudiar tot el que m'encuriosís de les diferents hemoglobines, ja que no hi havia un experiment establert i, per tant, era jo qui decidia què m'interessava.

De manera que el meu treball es pot esquematitzar en aquests objectius, els quals es veuran reflectits en els apartats entre parèntesis i en cursiva:

1. Entendre com es produeix l'evolució (*Com les mutacions porten l'evolució i l'especiació*)
2. Provocar l'evolució artificialment (*Evolució provocada de la proteïna HisF*)
3. Estudiar i comparar el resultat de l'evolució real (*Evolució i comparació de l'hemoglobina en diferents espècies*)

Les preguntes formulades abans de la realització del treball són, numerades segons l'apartat on es responen, les següents:

1. Què són les mutacions? I l'evolució? I l'especiació? Què les provoca?  
Quina relació hi ha entre els tres termes anteriors?
2. Podem provocar l'evolució artificialment?  
Mutacions introduïdes a l'atzar poden fer aparèixer noves proteïnes amb funció pròpia?  
Com podem demostrar-ho?
3. Una mateixa proteïna pot ser diferent en diverses espècies? Si és que sí, a on es reflecteixen aquests canvis?  
Quines seran més semblants entre elles?  
Podem establir relacions evolutives entre les espècies comparant les seves respectives proteïnes?  
Té alguna explicació evolutiva la situació de les zones menys o més conservades de les diverses proteïnes?  
Tindran propietats de flexibilitat diferent les proteïnes segons els canvis apareguts al llarg de l'evolució?

# Com les mutacions porten l'evolució i l'especiació

---

## Introducció

Les mutacions són normalment considerades una font de malaltia per la població i també pels mitjans de comunicació. Tot i així, les mutacions permeten la variabilitat entre els diferents genomes d'una espècie, cosa que significa que aquesta espècie està més capacitada per sobreviure en cas de canvis ambientals.

L'any passat, el 2009, pot ser considerat un dels més importants aniversaris de la ciència. Aquest any marca el 200è aniversari del naixement de Charles Darwin (12 de febrer del 1809) i el 150è aniversari de la publicació del seu llibre *On the Origin of Species by Means of Natural Selection* (1859). La publicació d'aquest llibre va suposar un canvi de visió sobre l'evolució de la vida i va revolucionar el concepte de l'aparició de la vida. Tot i que ell sabia que l'evolució de la vida es basava en la variabilitat, mai va arribar a conèixer que les mutacions n'eren les causants.

Més tard, els científics van descobrir que les mutacions eren una font de canvis genètics. La combinació d'aquesta descoberta amb l'antic Darwinisme va donar lloc a l'inici del Neodarwinisme, el qual postula que la selecció natural contribueix a la selecció positiva de les mutacions que donen un avantatge adaptatiu. Però no tothom accepta aquest fet; la teoria neutral argumenta que les mutacions no estan afectades per processos selectius, sinó que s'acumulen en les poblacions, les quals tenen la deriva genètica com a principal força evolutiva. Tot i així, ambdues teories accepten que les mutacions són la causa de l'evolució i la causa de tenir un gran nombre d'espècies actualment, fet que en fa important l'estudi.

## Mutacions

El genoma dels organismes no és estàtic; canvia gràcies a les mutacions, les quals són canvis en la seqüència del genoma.

És important classificar les mutacions depenent del tipus de cèl·lula on tenen lloc. Si una mutació ocorre en un gàmeta, passarà a la descendència, fet que significarà que cada cèl·lula del futur embrió portarà la mutació. En canvi, si una mutació no té lloc a un gàmeta, o sigui, una mutació somàtica, aquesta no es transmetrà a la descendència, sinó que només afectarà l'individu on es va originar la mutació i, per tant, no es fixarà a la població. Les teories evolutives se centren en els canvis d'ADN produïts als gàmetes perquè són aquestes mutacions les úniques que contribueixen a l'evolució.

També podem classificar les mutacions depenent de la direcció de transferència. Les mutacions de transferència vertical tenen lloc quan passen d'avantpassats a descendents. Les de transferència horitzontal ocorren quan l'ADN es mou d'un organisme a un altre. Aquestes mutacions es propaguen molt ràpid, ja que no hi ha diferència generacional. Un exemple és l'intercanvi de plasmidis entre bacteris. Per últim, les de transferència dins d'un organisme es donen quan els gens o part dels gens es mouen dins un organisme.

## Tipus de mutacions segons l'alteració dels nucleòtids

La classificació segons l'alteració dels nucleòtids és molt important, ja que ens permet ser conscients dels efectes que tindrà un cert tipus de mutació.

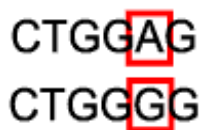
### Mutacions gèniques

Les mutacions gèniques són alteracions en la seqüència de nucleòtids d'un gen. Per poder entendre les conseqüències d'aquestes mutacions s'ha de tenir clar el procés de traducció i, sobretot, de transcripció:

Alguns gens codifiquen per proteïnes. Els gens codificadors de proteïnes contenen la informació sobre la seqüència d'aminoàcids de la proteïna a codificar. La informació es fa arribar als ribosomes, productors de proteïnes, mitjançant l'ARN. Cada grup de tres nucleòtids de l'ARN (codons) determina un aminoàcid de la proteïna (Imatge 2). El procés s'atura quan el ribosoma llegeix un codó d'acabament (stop codon). Un canvi en l'ADN provoca forçosament un canvi en l'ARN, i aquest pot provocar un canvi en els aminoàcids de la proteïna, fet que pot comportar un mal funcionament de la proteïna resultant.

Podem fer aquesta classificació:

- *Mutacions puntuals* (Imatge 1): Només un sol parell de bases és canviat per un altre parell de bases. Són l'error de duplicació més comú. Tot i ser un canvi petit, aquest pot comportar grans problemes de funcionament fisiològic, com en el cas de l'anèmia falciforme: tant l'hemoglobina normal com la falciforme tenen 147 aminoàcids. Ara bé, la falciforme té una valina (GTG) al sisè aminoàcid, enlloc d'un àcid glutàmic (GAG). Només el canvi de T a A fa canviar l'aminoàcid, fet que provoca que l'hemoglobina falciforme no transporti l'oxigen tan bé com l'hemoglobina normal, i també que tendeixi a bloquejar els capil·lars.



Imatge 1: Mutació puntual (Font: <http://evolution.berkeley.edu/evosite/evohome.html>)

Per sort, el nostre codi genètic (Imatge 2) està degenerat. Això significa que un aminoàcid pot estar codificat per diversos codons a la vegada, tres de mitjana. L'evidència és clara quan veiem que hi ha 64 codons possibles però només 20 aminoàcids diferents. Per exemple, AGU codifica per una serina; si el codó canviés a causa d'una mutació per AGC, tindríem una serina igualment. Així, una mutació puntual no sempre produirà un canvi en la seqüència d'aminoàcids i, per tant, no comportarà cap conseqüència (mutació silenciosa o neutral). En canvi, les mutacions que sí que afecten la proteïna s'anomenen mutacions no sinònimes.

		Second nucleotide					
		U	C	A	G		
First nucleotide	U	UUU Phe UUC UUA Leu UUG	UCU Ser UCC UCA UCG	UAU Tyr UAC UAA STOP UAG STOP	UGU Cys UGC UGA STOP UGG Trp	U C A G	Third nucleotide
	C	CUU Leu CUC CUA CUG	CCU Pro CCC CCA CCG	CAU His CAC CAA Gln CAG	CGU Arg CGC CGA CGG	U C A G	
	A	AUU Ile AUC AUA AUG Met	ACU Thr ACC ACA ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U C A G	
	G	GUU Val GUC GUA GUG	GCU Ala GCC GCA GCG	GAU Asp GAC GAA Glu GAG	GGU Gly GGC GGA GGG	U C A G	

Imatge 2: Codi genètic (Font: <http://www.nature.com/scitable>)

Quan una mutació puntual és comuna en la població utilitzem el terme SNP, "single nucleotide polymorphism". Tenim polimorfisme en un gen quan podem trobar com a mínim dues seqüències diferents per aquest gen, on cada seqüència ha d'estar representada com a mínim en l'1% de la població. Així, un polimorfisme ve d'una mutació que s'ha estès en un 1% o més de les persones en l'actualitat. De mitjana, podem trobar SNPs cada 1.000-2.000 nucleòtids del genoma humà (Clancy, 2008a).

- *Insercions* (Imatge 3) i *delecions* (Imatge 4), les quals s'agrupen amb el nom d'"índel": un o més nucleòtids són afegits o eliminats durant la duplicació de l'ADN. Si l'índel té lloc dins un exó d'un gen, podria afectar el pas d'ARN a proteïna a partir del punt de la mutació, ja que com hem dit anteriorment, cada tres nucleòtids codifiquen un aminoàcid. Si un o més nucleòtids s'afegeixen o s'eliminen, aquests canviaran tots els grups de tres nucleòtids previstos, fent que tots els aminoàcids canviïn o que aparegui un codó d'acabament abans del compte, fet que comportaria que la proteïna fos més curta que la proteïna correcta. Totes aquestes mutacions reben el nom de mutacions de desplaçament del marc de lectura. Així, només si l'índel consta d'un múltiple de tres nucleòtids, el canvi pot no ser important.

CTGGAG CTGGAG  
CTGGTGGAG CTAG

Imatge 3: Inserció Imatge 4: Deleció

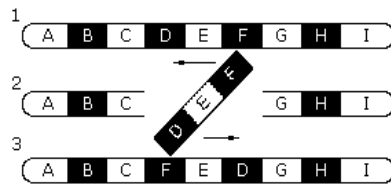
(Font: <http://evolution.berkeley.edu/evosite/evohome.html>)

### Mutacions cromosòmiques

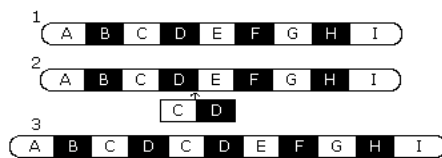
Són les mutacions que provoquen canvis en l'estructura interna dels cromosomes. De vegades poden afectar grans regions del genoma. Podem trobar:

- *Inversions* (Imatge 5): Es produeixen quan una regió d'un cromosoma se separa de la resta del cromosoma, gira 180° i es torna a col·locar al mateix lloc. No acostumen a perjudicar l'individu, ja que continua tenint el gen. Ara bé, si durant la meiosi es produeix un encreuament en aquesta zona, el descendent no tindrà el gen complet.

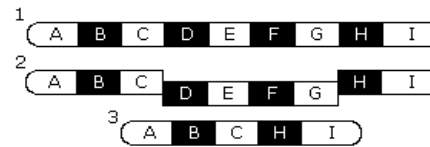
- *Delecions* (Imatge 6): Tenen lloc quan una regió del cromosoma es perd, fet que comporta l'absència de tots els gens del fragment. Si afecta els dos cromosomes homòlegs sol ser fatal, ja que l'individu passa a no tenir cap còpia de certs gens.
- *Duplicacions* (Imatge 7): Ocorren quan una regió d'un cromosoma es repeteix. Aquesta regió no necessàriament s'adhereix al cromosoma original. Com veurem més endavant, aquestes mutacions tenen importància per a l'evolució.
- *Translocacions* (Imatge 8): Es produeixen quan una regió d'un cromosoma s'enganxa a un altre cromosoma. L'individu continua tenint tots els gens i, per tant, no li suposarà cap problema. Ara bé, quan es reparteixin els cromosomes per a la descendència, aquesta tindrà gens repetits o li faltaran certs gens.



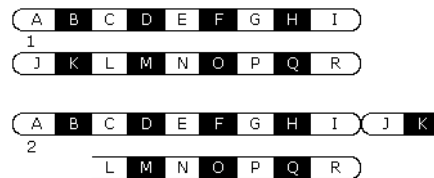
Imatge 5: Inversió (Font: <http://www.biology-online.org>)



Imatge 7: Duplicació (Font: <http://www.biology-online.org>)



Imatge 6: Deleció (Font: <http://www.biology-online.org>)



Imatge 8: Translocació (Font: <http://www.biology-online.org>)

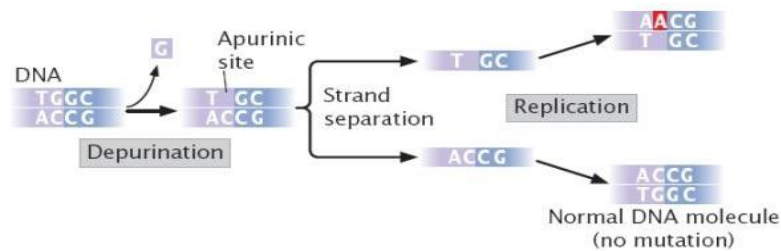
### Mutacions genòmiques

Són les alteracions en el nombre de cromosomes propi d'una espècie. Podem distingir:

- *Aneuploïdia*: És l'alteració en el nombre normal (generalment dos) d'exemplars d'un o més cromosomes, sense arribar a afectar el joc complet. Podem trobar nul·lusomies (cap exemplar d'un cromosoma concret), monosomies (només un exemplar), trisomies (tres exemplars),... Un exemple de monosomia és la síndrome de Turner, on les dones només tenen un cromosoma X i, per tant, un total de 45 cromosomes. Un cas molt conegut de trisomia és la síndrome de Down, on els afectats tenen tres exemplars del cromosoma 21, o sigui, un total de 47 cromosomes.
- *Euploïdia*: És l'alteració en el nombre normal de jocs de cromosomes d'un individu. Trobem monoploïdia quan l'individu només té un exemplar de cada tipus de cromosoma, i poliploïdia quan l'individu té més de dos exemplars de cada cromosoma. En els humans, tenir una monoploïdia significaria tenir 23 cromosomes, i una poliploïdia tenir-ne 69, 102, 125,... Cap d'aquests casos és viable en humans, ni en els respectius casos en la majoria d'animals, però és comú en plantes.

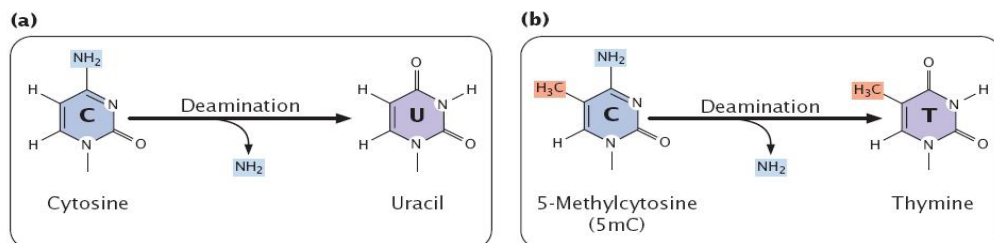
## Causes de l'aparició de mutacions

- *Canvis tautomèrics*: Cada base nitrogenada té dues formes diferents anomenades tautòmers. Es diferencien entre elles per les posicions de certs hidrògens. Una és considerada la normal (més freqüent), i l'altra és la rara. Les dues estan en equilibri i espontàniament passen de l'una a l'altra, produint canvis tautomèrics. Si aquest canvi succeeix durant la duplicació de l'ADN, s'introduiran mutacions en la nova cadena, ja que els tautòmers s'aparellen diferent: la forma rara de la G es complementa amb la T i també el cas invers, i la forma rara de l'A es complementa amb la C i el cas invers.
- *Mutacions espontànies*: Se'n coneixen dues classes. La primera és la despurinació (Imatge 10), on una G o una A es perd a causa d'un procés d'hidròlisi. Això pot provocar que s'incorpori un nucleòtid incorrecte durant la següent duplicació, ja que, per defecte, s'acostuma a afegir una adenina. Aquesta reacció és molt habitual; cada dia l'ADN de cada cèl·lula humana perd aproximadament 500 purines (Strachan, 1999).



Imatge 10: Despurinació (Font: Clancy, 2008a)

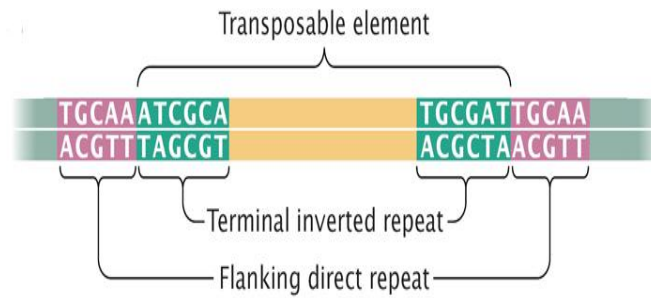
L'altre cas és el de la desaminació (Imatge 11). En aquesta mutació espontània, un grup amino del quart carboni d'una base nitrogenada és substituït per un àtom d'oxigen. Així, una citosina passa a ser un uracil, el qual no és una base nitrogenada de l'ADN, sinó de l'ARN (Imatge 11a). Els mecanismes de reparació detecten aquesta incongruència i tornen a canviar-ho per una citosina. Ara bé, si la desaminació es produeix en la forma tautomèrica de la citosina, obtindrem una timina (Imatge 11b), una de les bases nitrogenades de l'ADN i, per tant, no se substituirà. A partir d'aquí ja tindrem introduïda una mutació.



Imatge 11: Desaminació (Font: Clancy, 2008a)

- *Transposicions* (Imatge 12): són seqüències d'ADN que es mouen d'una posició del genoma a una altra, d'aquí que s'anomenin "jumping genes". En general s'anomenen elements genètics transposables (TEs, en anglès), i poden ser més petits que un gen (seqüències d'inserció), un gen o un grup de gens (transposons).

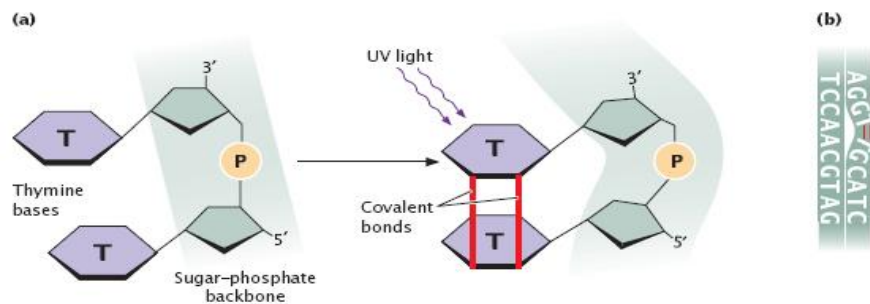
Es van identificar per primer cop fa més de 50 anys per la genetista Barbara McClintock del Cold Spring Harbor Laboratory a New York (McClintock, 1950, 1953). Els TEs es troben en gairebé tots els organismes i normalment en gran nombre. Per exemple, els TEs fan aproximadament el 50% del genoma humà i el 90% del de blat de moro (Pray, 2008a).



Imatge 12: Estructura d'un TE (Font: <http://www.nature.com/scitable>)

Si un TE se situa dins d'un gen, pot modificar el marc de lectura del gen o bé interrompre la seva expressió. Si se situa fora d'un gen pot provocar mutacions cromosòmiques. Tenen una gran habilitat d'incrementar la diversitat genètica i això els converteix en una part important de l'evolució en tots els organismes que porten aquestes seqüències. Per exemple, en la *Drosophila*, el 50-85% de les mutacions estan causades per TEs (Pray, 2008a).

- *L'ambient*: Existeixen dos tipus de bases nitrogenades en l'ADN: les purines (adenina i guanina) i les pirimidines (timina i citosina). Els rajos ultraviolats (UV) poden causar dímers entre pirimidines (C-C, C-T o T-T) (Imatge 13), els quals són enllaços covalents entre dues pirimidines que estan de costat en una cadena d'ADN (Imatge 13a). Aquests enllaços distorsionen la forma de la cadena (Imatge 13b). Tot i que existeix un mecanisme per a desfer-se d'aquests dímers, sovint no s'arreglen. Això fa que el següent cop que es copï l'ADN, només es col·loqui un nucleòtid com a complementari del dímer. Per aquest motiu, els rajos UV poden causar càncers de pell.



Imatge 13: Dímer de pirimidina (Font: Clancy, 2008b)

A més, els organismes també estem exposats a altres radiacions de ionització més energètiques com ara els rajos còsmics, els rajos gamma i els rajos X. Aquests rajos trenquen la doble cadena d'ADN i acostumen a comportar mutacions quan els sistemes de reparació volen arreglar el mal produït. A diferència dels rajos UV, aquestes formes de radiació poden penetrar el teixit epitelial, de manera que poden causar mutacions a qualsevol lloc de l'organisme.

- *Productes químics*: Certs productes químics tenen la capacitat de modificar els nucleòtids, i d'aquesta manera provoquen mals aparellaments de bases. Altres es poden col·locar entre els parells de base de l'ADN, propiciant l'aparició de deletions i insercions.
- *Fusió i fissió cèntrica*: En la fusió, dos cromosomes no homòlegs s'uneixen, donant lloc a un sol cromosoma. La fissió és el cas invers: un cromosoma es divideix en dos.



- *Errors durant la meiosi:* De vegades, durant la meiosi hi ha una distribució errònia de les cromàtides homòlogues, fet que resulta amb una cèl·lula amb les dues cromàtides i amb l'altra cèl·lula sense cap cromàtide. Aquest error juntament amb la fusió i fissió cèntrica són els causants de les aneuploidies. Si, després la meiosi, obtenim gàmetes  $2n$  enlloc de  $n$ , podem obtenir poliploidies si aquests gàmetes en fecunden de normals o també  $2n$ . A més, si l'encreuament durant la meiosi no es duu a terme correctament, pot provocar certes mutacions cromosòmiques.
- També hi ha mutacions causades per la habilitat de certs virus i bacteris. Els virus poden inserir una còpia d'ells mateixos dins del genoma d'un altre organisme, i els bacteris poden intercanviar plasmidis, els quals són petites fragments d'ADN circulars.

## Freqüència de mutacions

La freqüència de mutacions varia considerablement depenent de l'organisme i també depenent de la part del genoma a estudiar. La freqüència de mutacions més baixa s'ha trobat en bacteris amb valors d'un error entre 100 milions i un bilió de nucleòtids, mentre que la més alta s'ha trobat en humans en una freqüència d'un error entre 100 i 1.000 nucleòtids (Pray, 2008b).

És prou acceptat actualment que les mutacions són produïdes a l'atzar en l'ADN. Per tant, hi hauria d'haver el mateix nombre de mutacions en qualsevol part del genoma. Ara bé, la freqüència de mutacions que s'han fixat sí que depèn de la regió del genoma. Per exemple, trobem una freqüència de mutacions inferior en els exons (ADN codificador), ja que les mutacions en aquesta part del genoma acostumen a ser perjudicials o letals, de manera que són seleccionades negativament. Els promotors, que són ADN no codificant situats abans del gen, i els introns, que són ADN no codificant situats dins del gen, tenen també una freqüència de mutacions inferior que la resta de l'ADN no codificant perquè tenen un paper en la regulació de l'expressió del gen. És per aquest motiu que els gens i els promotors tenen una preservació evolutiva elevada.

El principal problema en l'estimació correcta de la freqüència de mutacions és que l'ADN té sistemes de reparació de l'ADN, fet que no ens permet detectar totes les mutacions produïdes. Per poder calcular la freqüència de mutacions necessitem tenir una seqüència d'ADN comuna pels descendents que estudiem. Llavors se seqüencia aquest mateix fragment en els descendents i es conten totes les mutacions observables. Finalment, aquestes observacions es combinen amb el nombre de generacions que connecten els individus estudiants per tal de calcular la freqüència de mutacions total.

Les mutacions poden ser beneficioses, perjudicials o neutrals depenent del lloc del genoma on tenen lloc i de l'ambient on viu l'organisme que porta la mutació. Per exemple, només una mutació que faci un organisme resistent a un producte serà útil si l'organisme viu en un ambient amb aquest producte. És molt important recordar que els factors ambientals poden afectar la freqüència de mutacions però no la direcció de les mutacions. Per exemple, si un organisme està exposat a productes químics perjudicials, la freqüència de mutacions potser augmentarà, però en cap cas aquestes mutacions faran necessàriament l'organisme resistent a aquests productes químics. És per això que es diu que les mutacions són a l'atzar i sense propòsit.

## Evolució

Per a tractar l'evolució hem de tenir presents certs mots, els quals apareixen en aquest paràgraf. Existeixen diferents formes d'un mateix gen, les quals anomenem **al·lels**, i que són productes de mutacions en aquest gen. Aquest fet és bàsic per estudiar la genètica de població. Una **població** és un grup de membres de la mateixa espècie que s'encreuen entre ells. Així, la genètica de població estudia la variació que ocorre entre els gens d'una població. El **patrimoni genètic** és la col·lecció de tots els al·lels diferents d'una població. La **variació genètica** d'una població es pot mesurar segons el nombre d'al·lels diferents i la seva freqüència. Aquesta variació és alta quan hi ha molts al·lels diferents en tots els gens i quan hi ha moltes combinacions diferents d'aquests al·lels. La variació genètica és dinàmica, o sigui, està canviant constantment, de manera que les freqüències genètiques varien al llarg del temps (Definicions de Scitable).

L'evolució és un procés que resulta dels canvis en el material genètic d'una població al llarg del temps. Si no hi haguessin mutacions, tots els individus tindrien el mateix material genètic, fet que faria impossible l'existència de canvis en el material genètic. Per tant, les mutacions són crucials per a l'evolució.

### Sistemes que provoquen la variació del patrimoni genètic

Tal com Ulrich Kutschera (2009) diu en article sobre Darwin, "Avui sabem que la reproducció sexual, a través de la recombinació genètica, combinada amb les mutacions hereditàries, és el procés clau per aconseguir variabilitat entre poblacions d'animals o plantes".

Gràcies a les mutacions obtenim nous al·lels. A més, gràcies a mutacions també poden aparèixer nous gens: si primer tenim un gen doblat, ja sigui per una duplicació cromosòmica, per TEs o per poliploidia, i un dels gens muta, continuarem tenint un gen que durà a terme la funció normal d'aquest més un altre gen que pot dur a terme una nova funció. Aquest increment de gens sembla que ha tingut un paper important en l'evolució. Es pensa que les primeres criatures tenien un nombre de gens mínim, els essencials per a sobreviure i reproduir-se. Per tant, l'augment de la llargada del genoma ha permès la diversificació genètica.

La reproducció sexual és la producció de descendència la reserva genètica de la qual és una barreja de la reserva de dos gàmetes diferents (Definició de Scitable). Aquest procés és molt més important que les mutacions en els éssers vius que la duen a terme, ja que consta de tres passos on cada un ajuda a fer que l'individu resultant sigui realment únic:

- *Distribució independent dels cromosomes*: Quan s'acaba el procés de meiosi I tenim dues cèl·lules haploides. Les cèl·lules haploides resulten de la distribució d'un cromosoma homòleg per cada cèl·lula. Aquest repartiment és independent per cada parell de cromosomes.
- *Encreuament*: El trobem al principi de la profase I. Els dos cromosomes homòlegs es posen de costat, de manera que una cromàtide d'un està alineada nucleòtid per nucleòtid amb una cromàtide de l'altre. L'ADN d'aquestes dues cromàtides es trenca al mateix lloc i després es torna a unir a la cromàtide de l'homòleg. Això permet barrejar el material genètic del mascle i de la femella en el mateix cromosoma, i així obtenim que totes les cromàtides dels dos cromosomes homòlegs siguin diferents entre elles.

- **Fertilització a l'atzar:** Resulta de la combinació a l'atzar de dues cèl·lules haploides. Per exemple, en els humans, tant un òvul com un espermatozoide representen un possibilitat entre 8 milions. Així, una combinació a l'atzar entre un òvul i un espermatozoide representa una possibilitat entre 64 milions (Campbell et al., 2007).

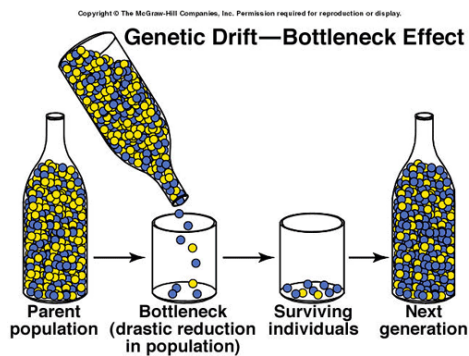
Tot i que les mutacions i la reproducció sexual són els factors que creen la variabilitat, no són els principals factors que alteren la freqüència al·lèlica, ja que la reproducció sexual només barreja els al·lells però no en canvia la freqüència, i les mutacions, tot i que poden modificar la freqüència al·lèlica, els canvis d'una generació a l'altra són realment petits. De manera que els tres factors principals que alteren la freqüència al·lèlica i produeixen la major part de les modificacions evolutives són la selecció natural, la deriva genètica i el flux gènic.

- **Selecció natural:** Darwin va ser el principal introductor d'aquesta teoria. Kutschera (2009) també cita frases del llibre *Origin of species* de Darwin al seu article, que descriuen la selecció natural com a "la preservació de les variacions favorables i el rebuig cap a les variants perjudicials". Després, una altra que especifica "Es pot dir que la selecció natural està inspeccionant cada dia i cada hora totes les variacions...; rebutjant aquella que és dolenta, preservant i acumulant totes aquelles que siguin bones". Avise (2009) també cita una frase de Darwin en un article que diu "Neixen més individus en una espècie dels que poden sobreviure. Conseqüentment, hi ha una lluita freqüent per a l'existència, de manera que si un ésser varia d'una manera beneficiosa per ell mateix, sota les complexes i de vegades variables condicions de vida, tindrà una millor oportunitat per a sobreviure i, per tant, serà seleccionat naturalment. A partir del fort principi de l'herència, qualsevol varietat seleccionada tendirà a propagar la seva forma nova i modificada".

La majoria de les mutacions no sinònimes redueixen l'eficiència de l'organisme. Aquestes són seleccionades negativament i desapareixen de la població: selecció negativa o purificadora. Quan una mutació no afecta l'organisme, passa desapercibuda per la selecció. Finalment, quan una mutació incrementa la possibilitat de l'organisme a tenir més descendents, és subjecte de la selecció positiva o avantatjosa, de manera que es podrà escampar entre la població.

Com que el paper de les condicions variables de l'ambient no era molt precís, la teoria de la selecció natural de Darwin va ser qüestionada per molts científics durant les primeres dècades del segle XX. Tots els autors acceptaven el principi de Malthus: cada espècie tendeix a créixer en una freqüència ràpida i produeix més individus dels que poden sobreviure. A més a més, ells estaven d'acord en el fet que aproximadament el 98% d'aquests individus moren abans d'arribar a la maduresa. Només aproximadament el 2% sobreviuen després de la lluita per l'existència per a produir la següent generació. Tot i així, es discutia que no era clar si realment els individus que sobreviuen eren els que posseïen variacions que els permetien ser més eficients que els altres. El 1918, Tower va concloure que encara no ha estat demostrat que els individus que sobreviuen eren els possessors afortunats de modificacions avantatjoses, sinó que ell creia que era més una qüestió de la posició casual dels individus en el moment que tenien lloc els "accidents de la vida" (Kutschera, 2009).

- **Deriva genètica:** Descriu les fluctuacions a l'atzar en el nombre d'al·lels en una població. En una població petita, la deriva genètica pot ocórrer quan els gàmetes de la següent generació no contenen tots els al·lels de la població anterior o quan els contenen en una freqüència diferent. Però també la podem trobar causada per altres fenòmens, com l'efecte coll d'ampolla i l'efecte fundador. L'efecte coll d'ampolla (Imatge 14) succeeix quan un desastre causat per un canvi sobtat en el medi redueix dràsticament la mida de la població. L'efecte fundador ocorre quan pocs individus se separen de la població principal i formen una nova població. En tots dos casos, el patrimoni genètic dels supervivents o dels individus aïllats no serà necessàriament el mateix que el de la població original. La deriva genètica és més pronunciada en petites poblacions perquè tenen menys variabilitat i, per tant, una habilitat inferior per adaptar-se a condicions canviants. Per aquesta raó, com més variabilitat hi ha en una població, més preparada estarà la població a adaptar-se a canvis.



Imatge 14: Efecte ampolla (Font: [www.biology.unm.edu/unmbio/index.shtml](http://www.biology.unm.edu/unmbio/index.shtml))

- **Flux gènic:** És el canvi de la freqüència al·lèlica que es produeix quan diferents poblacions es barregen. Aquest intercanvi genètic redueix les diferències entre les poblacions i, si el contacte és prou llarg, les dues poblacions poden passar a ser una única població amb un únic patrimoni genètic. Ara els humans ens movem per tot el món molt més que al passat, de manera que el flux gènic ha esdevingut un important agent de canvi evolutiu entre les poblacions d'humans que abans estaven aïllades.

## Especiació

Un punt essencial en la teoria de l'evolució és l'origen de noves espècies, que poden ser definides com a “grups de poblacions naturals que estan aïllades reproductivament d'altres grups de poblacions naturals” (Mayr, 1963). Hi ha dues classes de canvi evolutiu depenent del nombre d'espècies resultant. Un és l'anagènesi, on els canvis heretats alteren les característiques de tota l'espècie i, quan hi ha suficient diferència entre l'espècie original i la nova, es pot donar un nom nou a l'espècie actual. L'altre és la cladogènesi, on la reserva genètica se separa en dues o més reserves i cada una dona lloc a una nova espècie. Només la cladogènesi provoca varietat biològica ja que fa augmentar el nombre d'espècies.

L'especiació es pot veure com “el procés pel qual dues poblacions idèntiques divergeixen genèticament fins al punt que és impossible el seu encreuament altre cop. Les espècies són, per tant, genèticament diferents i independents” (Wu et al., 2004). L'especiació es produeix quan apareix l'aïllament reproductiu (RI, en anglès), el qual es pot definir com el no intercanvi de gens entre dues espècies que estan en contacte entre elles (Wu et al., 2004), o bé com les barreres biològiques que eviten que dues espècies tinguin híbrids fèrtils i viables (Campbell et al. 2007).

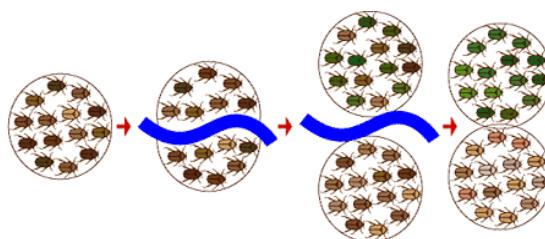
La RI pot ser causada per dues classes de mecanismes. En el primer, les dues poblacions no s'aparellen o eviten la fertilització, ja sigui perquè viuen en llocs separats, tenen rituals d'aparellament diferents, tenen aparença diferent o perquè els gàmetes són incapaços de fertilitzar l'altra població. L'altre mecanisme el trobem després de l'aparellament, on les dues espècies s'aparellen però els híbrids són inviables, estèrils o mal adaptats a l'ambient.

Es creu que hi ha uns gens responsables d'aquest aïllament reproductiu, els quals s'han passat a anomenar “gens d'especiació” (Wu et al., 2004). Les espècies de *Drosophila* són les més utilitzades en aquest estudi i, després de recopilar un gran nombre d'informació, es pensa que cada gen d'especiació té poc efecte per si sol però que, en combinació, aquests gens causen incompatibilitat reproductiva. També es duen a terme estudis per a trobar la funció usual dels gens que produeixen el RI després de l'aparellament, ja que no s'accepta com a possible que la seva única funció sigui la d'esterilitzar la descendència.

### Tipus d'especiació

L'especiació pot actuar per dos camins diferents depenent de la manera que el flux genètic és interromput:

- **Especiació al·lopàtrica** (Imatge 15): Les poblacions divergents són geogràficament separades sense que hi hagi flux genètic entre elles. Aquest aïllament es pot produir quan hi ha canvis geogràfics que divideixen la població o bé quan una part de la població conquereix una nova terra i resta reparada de la població original. Per a alguns individus aquestes barreres geogràfiques no són un problema, com en els ocells o les plantes que duen a terme la pol·linització utilitzant el vent per endur-se les llavors o el pol·len. Altres animals, en canvi, queden totalment aïllats, com els petits rosegadors.



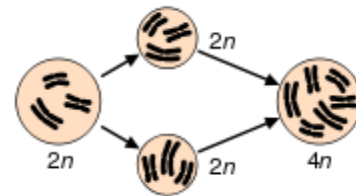
Imatge 15: Especiació al·lopàtrica (Font:

<http://evolution.berkeley.edu/evosite/evohome.html>)

Un cop aquesta separació s'ha produït, les reserves genètiques de les dues poblacions es van separant gràcies a mutacions diferents, pressions selectives diferents i derives genètiques, factors que canviaran la freqüència genètica i que potser produiran RI. Només podem parlar que s'ha produït aquesta especiació si les poblacions han canviat tant que ja no es poden aparellar o no poden donar descendència fèrtil.

Tot i que hi ha unes barreres que eviten el contacte entre les poblacions, és important tenir en compte que són les diferències genètiques entre les poblacions les que interrompen el flux genètic, i no les barreres per si soles.

- **Especiació simpàtrica:** Les poblacions no se separen geogràficament. Certs canvis cromosòmics i encreuaments no aleatoris són els que redueixen el flux genètic i els que, per tant, són els responsables de l'especiació. L'especiació cromosòmica acostuma a tenir lloc en plantes, i la manera més comuna d'especiació és mitjançant la poliploïdia (Imatge 16), en la qual el mida del genoma es duplica, de manera que l'organisme té dos jocs del mateix genoma (autotetraploïdia) o té dos jocs diferents de genomes (alotetraploïdia). La primera succeeix a causa d'una mala divisió cel·lular, on els gàmetes no passen a ser  $n$ . L'organisme resultant no pot encreuar-se amb la generació anterior, però aquesta planta pot autopol·linitzar-se o bé creuar-se amb altres plantes tetraploides. El segon cas té lloc quan dues espècies diferents es creuen i produeixen un híbrid. Aquest híbrid és infèrtil, però es pot escampar mitjançant la reproducció asexual, típica en plantes.



Imatge 16: Poliploïdia (Font: <[www.newworldencyclopedia.org](http://www.newworldencyclopedia.org)>)

Pel que fa als encreuaments no aleatoris, podem trobar diferents possibilitats tant en animals com en plantes. Un cas el podem trobar quan factors genètics permeten a una subpoblació explotar una font d'aliment diferent de la de la població original (Imatge 17). Llavors es troba que els individus tendeixen a preferir els individus que s'alimenten de la mateixa font, provocant una separació invisible entre els dos grups. Un altre cas és quan una població és polimòrfica i algunes femelles només volen aparellar-se amb els mascles que tenen un polimorfisme específic i les altres femelles amb els mascles que tenen l'altre polimorfisme. En tots dos casos les poblacions poden acabar donant lloc a dues poblacions i, al final, esdevenir dues espècies diferenciades si es produeix RI.



Imatge 17: Explotació de diferents fonts d'aliments per la subpoblació i la població original (Font: <<http://evolution.berkeley.edu/evosite/evohome.html>>)

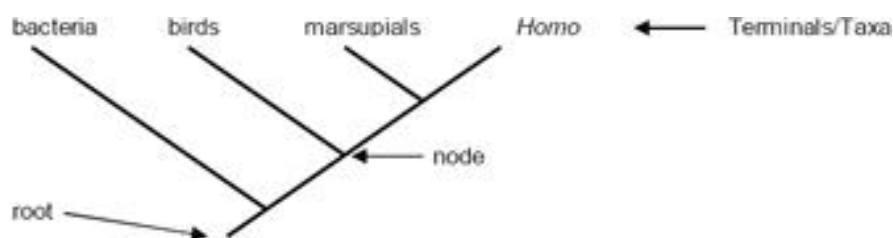
John C. Avise i Francisco J. Ayala (2009) van escriure en un seu article sobre Darwin que "Irònicament, l'*Origen de les espècies* diu relativament poc sobre l'evolució de les barreres que provoquen l'RI, les quals, sota el concepte biològic actual d'espècies, són clau per entendre els processos cladogènics (d'especiació)".

## Arbres filogenètics

Després que una espècie X pateixi un procés d'especiació i que les espècies resultants també en pateixin, podem obtenir un gran nombre d'espècies, les quals tindran un mateix avantpassat comú, l'espècie X. Arribats a aquest punt, el que potser ens pot interessar és saber com estan relacionades evolutivament totes aquestes espècies: quines estan més relacionades amb quines, quin és l'avantpassat més pròxim de dues d'elles, etc... Per tal d'obtenir aquest coneixement es realitzen arbres filogenètics.

Un arbre filogenètic és “un diagrama que representa les línies dels descendents evolutius de diferents espècies, organismes, o gens, a partir d'un avantpassat comú. Aquests arbres són útils per organitzar el coneixement existent sobre la diversitat biològica del nostre planeta, per estructurar classificacions i per revelar-nos fets que van passar durant l'evolució.” (Definició de Scitable).

Molts arbres filogenètics són arrelats. Això vol dir que hi ha una branca al principi que representa l'avantpassat comú de totes les espècies de l'arbre. Normalment no s'indica quin és aquest avantpassat. El que sí s'etiqueta són les puntes de l'arbre, on cada una representa una branca. Una punta pot representar un sol gen, un sol organisme, una espècie o un conjunt d'espècies. Els punts contraris a les puntes són els nodes, que representen el punt de naixement de les branques a causa de processos d'especiació. Per tant, són els últims avantpassats comú de les branques que se'n deriven.



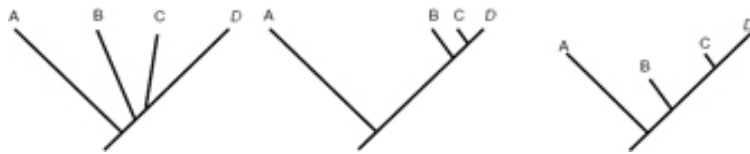
Imatge 18: Parts d'un arbre filogenètic (Font: Baum, 2008)

A l'hora d'analitzar un arbre filogenètic no és important en quina orientació està dibuixat l'arbre. Els tres arbres de la Imatge 19 ens aporten la mateixa informació perquè tenen la mateixa topologia, mot que significa de quines branques neixen altres branques. No s'ha de tenir en compte si les branques són diagonals, rectes o arrodonides.



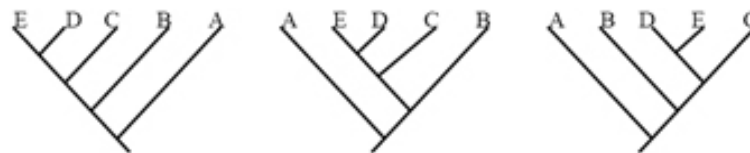
Imatge 19: Diferents tipus d'arbres filogenètics amb el mateix significat (Font: Baum, 2008)

Si no s'indica el contrari, un arbre filogenètic només descriu la història d'un avantpassat comú en forma de branques. Això vol dir que la llargada de les branques és irrellevant; simplement estan dibuixades amb una llargada o una altra per qüestions estètiques. Per això els arbres de la Imatge 20 contenen la mateixa informació. El que importa, com al cas anterior, és la topologia.



Imatge 20: Diferents arbres filogenètics amb el mateix significat  
(Font: Baum, 2008)

A part de la llargada i la forma de l'arbre, el que tampoc importa és l'ordre de les branques. Això significa que els nodes poden girar, de manera que la branca de la dreta passi a l'esquerra i a l'inrevés, com es veu a la Imatge 21. Ara bé, només podem canviar l'ordre de les branques sempre i quan una branca sempre surti del mateix node, ja que sinó canviaríem la informació que ens proporciona l'arbre.

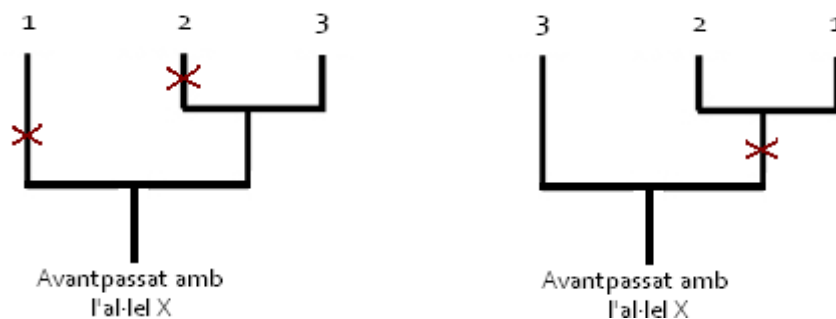


Imatge 21: Diferents arbres filogenètics amb el mateix significat  
(Font: Baum, 2008)

### Com obtenir un arbre filogenètic

Si es treballa amb poques seqüències curtes, l'arbre filogenètic es pot arribar a fer a mà. Però, com que aquests casos no acostumen a ser els més comuns, s'han arribat a idear molts mètodes bastant complexos que dibuixen arbres filogenètics automàticament; només cal inserir les seqüències d'ADN o de proteïnes que es vol que apareixin a l'arbre filogenètic.

Molts sistemes utilitzen el principi de mínima evolució i el principi de màxima parsimònia (Saitu et al., 1987). Imaginem-nos que tenim tres espècies on la 1 i la 2 tenen l'al·lel Y, mentre que la 3 té l'al·lel X del mateix gen. Totes tres provenen d'un avantpassat amb l'al·lel X. Podríem intentar explicar-ho amb aquests dos arbres filogenètics:



Les creus marcades en els arbres representen punts on es va produir una mutació de la qual en resulta l'aparició de l'al·lel Y, el qual substitueix l'al·lel X.

En tots dos casos suposem que l'avantpassat ha patit un procés d'especiació, donant lloc a dues espècies diferents. En el primer cas, una d'aquestes, després de sofrir una de les mutacions que provoquen el canvi de l'al·lel X per l'al·lel Y, ha acabat donant l'espècie 1. L'altra espècie resultat de l'especiació ha tornat a patir un procés d'especiació, donant lloc a l'espècie 3 i a una altra espècie que, després de sofrir una de les mutacions, ha acabat sent



l'espècie 2. D'aquesta manera, l'espècie 3 conserva el gen de l'avantpassat (l'al·lel X), però les altres dues han sofert la mateixa mutació i tenen l'al·lel Y.

En el segon cas, la primera especiació ha donat lloc a l'espècie 3 i a una altra espècie que patirà una de les mutacions. Aquesta segona, després de patir la mutació, pateix una especiació que resulta amb l'aparició de l'espècie 1 i 2. Així, l'espècie 3 conserva el gen de l'avantpassat (l'al·lel X) i l'espècie 1 i 2, com que provenen d'una mateixa espècie que tenia l'al·lel mutat, tenen totes dues l'al·lel Y.

El resultat de tots dos casos és exactament el mateix si només tenim en compte l'estudi del gen amb els al·lells X i Y. Per tant, en principi tots dos haurien de ser igual de vàlids i no podríem saber quin dels dos arbres és correcte. No obstant, hem de tenir clar que si n'haguéssim de triar un, agafaríem el segon sense dubtar-ho. El perquè és ben senzill: en el primer cas hem necessitat dues mutacions per aconseguir el mateix resultat que el segon cas, en el qual només s'ha produït una mutació. I, en termes evolutius, és sempre més probable el cas que necessiti menys evolució i, per tant, el que utilitzi més parsimònia.

D'aquesta manera, els mètodes que generen els arbres filogenètics estudien tots els possibles casos i trien el que millor segueix el principi de mínima evolució i el de màxima parsimònia. És el millor criteri per acostar-se a l'arbre real.

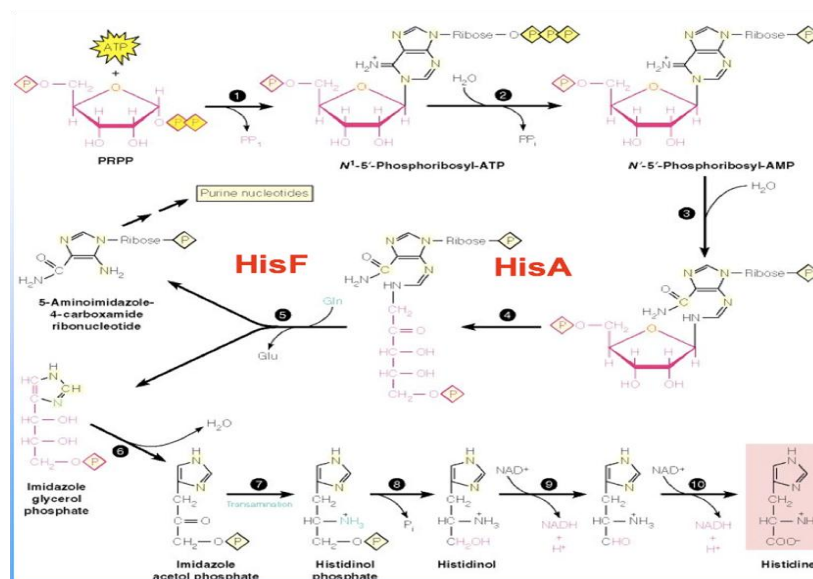
# Evolució artificial de la proteïna HisF

## Introducció

L'objectiu d'aquest apartat és intentar veure empíricament el funcionament de la creació de noves proteïnes: com a resultat de les mutacions en un gen, aquest passa a codificar una nova proteïna que pot ser que dugui a terme una funció diferent de la inicial.

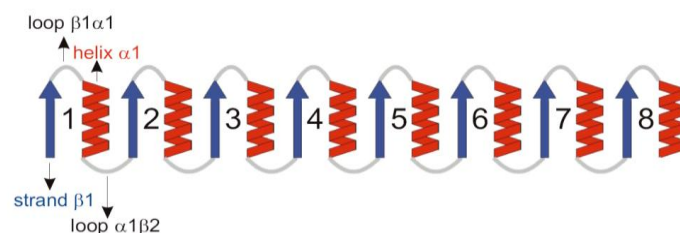
## HisA i HisF

Treballarem amb dues proteïnes, la HisA i la HisF. Són dos enzims que catalitzen dues reaccions successives en la biosíntesi de la Histidina (Imatge 1), un aminoàcid necessari per a la creació de proteïnes. El procés parteix de la glucosa, i no podrà acabar si no tenim un dels dos enzims, o sigui, que tots dos enzims són imprescindibles i igual d'importants.



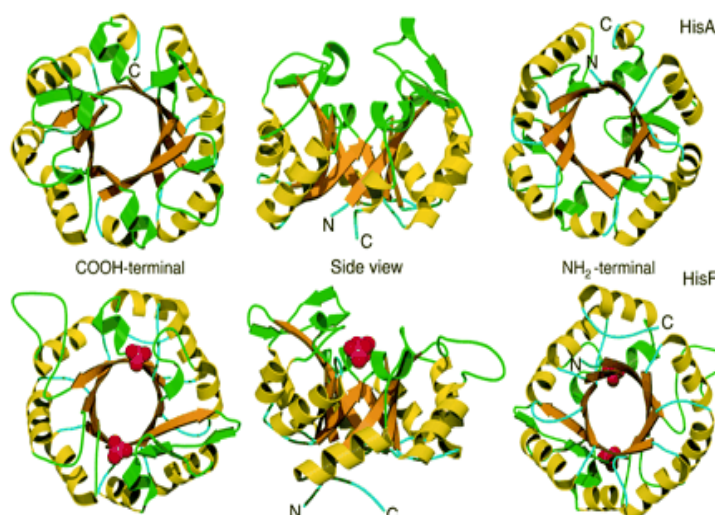
Imatge 1: Procés de biosíntesi de la Histidina, amb els punts d'acció de la HisA i la HisF marcats (Font: Henn-Sax, 2008)

Se sap que tots dos enzims tenen 250 aminoàcids, que 3 o 4 aminoàcids són els responsables de la funció de la proteïna i que la resta permeten l'estabilitat de la proteïna. Les dues proteïnes comparteixen el 25% de la seva seqüència (Lang et al., 2000) i són molt similars pel que fa a l'estructura: totes dues són proteïnes monomèriques, que vol dir que només tenen una cadena d'aminoàcids, i les seves estructures secundària i terciària és semblant. La seva estructura secundària (Imatge 2) és formada per una successió de làmines  $\beta$  i hèlix  $\alpha$ , les quals estan unides per zones sense plegar.



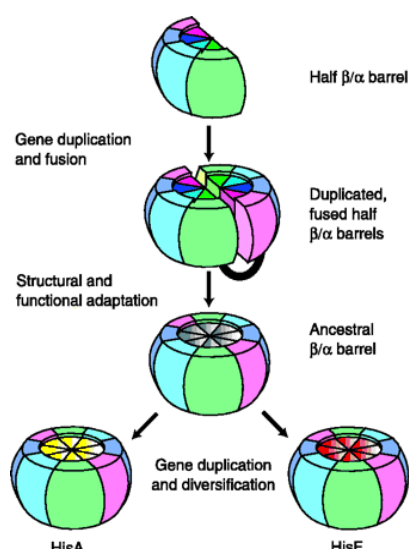
Imatge 2: Estructura secundària de la HisA i la HisF (Font: Henn-Sax, 2008)

A la Imatge 3 podem veure la seva estructura terciària o estructura atòmica. La diferència que més ressalta entre les dues proteïnes és que la HisF té dos ions fosfor (en vermell) enganxats a la cara activa.



Imatge 3: Estructures atòmiques de la HisA i la HisF en l'espècie *Thermotoga maritima* (Font: Lang et al., 2000)

Per aquestes semblances la majoria de científics creuen que, durant un període de temps del passat, els dos enzims eren en realitat un sol enzim. A causa de l'evolució, han acabat formant dos enzims diferents. El que es creu que va passar està representat en el model de la Imatge 4:



Imatge 4: Model de l'evolució de la proteïna que va donar lloc a la HisA i la HisF (Font: Lang et al., 2000)

El primer dibuix representa la proteïna original. Es pensa que el gen que la codificava va patir una duplicació gènica, la qual cosa va significar la generació de dues proteïnes idèntiques, les quals es van fusionar i van donar lloc al tercer dibuix. Una altra duplicació gènica seguida de la diversificació entre les dues còpies va donar lloc a dos enzims amb activitats catalítiques diferents, la HisA i la HisF.

Després de certs estudis, s'ha determinat que, en alguns casos, només canviant un aminoàcid d'una de les dues proteïnes s'aconsegueix que aquesta proteïna modificada faci la funció de l'altre. És precisament aquest canvi de funcionalitat mitjançant mutacions el que comprovem amb el següent experiment.

Nota important: Si escrivim HisA o HisF (la primera lletra en majúscula) ens estem referint a les proteïnes. En canvi, si escrivim hisA o hisF (la primera lletra en minúscula), ens estem referint als gens.

# Plantejament de l'experiment

## Pregunta

Podem modificar el gen *hisF* per tal d'aconseguir una proteïna que pugui fer la funció de la *HisA* en el procés de creació de la Histidina?

Nota: No ens proposem modificar la *HisA* per obtenir la *HisF* perquè, tal i com es pot veure en la Imatge 1, la reacció que catalitza l'enzim *HisA* és més fàcil que la que catalitza l'enzim *HisF*. Per això serà més senzill modificar el gen *hisF* perquè produeixi una proteïna funcional que hagi de dur a terme una reacció més simple. Tot i així, també és possible modificar el gen *hisA* perquè produeixi una proteïna funcional com la *HisF*.

## Hipòtesi

Utilitzant la error-prone-PCR podem mutar el gen *hisF* per tal que codifiqui per una proteïna la funció de la qual sigui equivalent a la *HisA*.

## Dedució

Si a un bacteri sense el gen *hisA* que es troba en un medi sense Histidina li introduïm un plàsmid amb el gen modificat de la *hisF* capaç de sintetitzar una proteïna amb la funció de la *HisA*, el bacteri sobreviurà.

Nota: Treballarem amb bacteris perquè:

1. Es reproduïxen molt ràpidament i, per tant, és més fàcil visualitzar-ne l'evolució.
2. És fàcil introduir-los gens nous mitjançant els plasmidis.
3. Una mutació és segur que passa a la descendència, no com en els humans que, perquè sigui així, ha de ser o l'òvul o l'espermatozoide que formaran el nou ésser.
4. Només tenen una còpia de cada gen, així que una mutació en aquesta còpia s'expressarà segur; en els humans, potser aquesta mutació passaria a donar un lloc a un al·lel recessiu i no s'expressaria.

## Disseny de l'experiment

Primer produïrem moltes còpies del gen *hisF* utilitzant el sistema de la PCR, tot i que farem que el resultat no siguin còpies idèntiques com de costum, sinó que introduïrem mutacions. D'aquesta manera obtindrem un gen que anomenarem *hisF\**.

Segonament aïllarem plasmidis d'una colònia de bacteris. Aquests seran els plasmidis on introduïrem el nostre producte de la PCR. Per fer-ho utilitzarem la restricció per tallar el plasmidi i per aïllar el gen *hisF\** del fragment de la PCR. Un cop fet això, podrem ajuntar-los mitjançant la lligació. D'aquesta manera ja tindrem preparats els nostres plasmidis amb el nostre gen modificat.

El que ara necessitem és introduir els plasmidis a bacteris, preparant una colònia de bacteris sense el gen *hisA* per ser químicament competents i aplicant la tècnica del "heat shock".

Per últim, prepararem dos tipus de medis pels bacteris: un amb tots els nutrients necessaris per la cèl·lula, i l'altre sense Histidina. Hi aplicarem els nostres bacteris i també altres que ens serviran de control. Per anar bé, els nostres bacteris haurien de poder créixer en el medi sense Histidina, ja que han de tenir tots els gens necessaris per a poder fabricar-se-la.

# Disseny de l'experiment

## Error-prone-PCR

El primer que volem és fer moltes còpies del gen hisF. Per això utilitzarem la PCR. Tot i així, no volem que totes les còpies siguin com el gen original. De fet, volem tot el contrari: si es produeixen mutacions durant el procés, potser aconseguirem que la seqüència de l'hisF doni lloc a una proteïna funcional com la HisA.

Per això, més que fer una PCR farem una error-prone-PCR, que és el mateix sistema que la PCR però amb alguns canvis. Per saber com fer una PCR hem de conèixer com es duplica l'ADN normalment.

### Duplicació de l'ADN

Per a la duplicació de l'ADN es necessiten més de 12 enzims. Les dues cadenes d'ADN de la doble hèlix serveixen com a patró per a fer les noves còpies d'ADN. Primerament, les dues cadenes es desenrotllen gràcies a l'enzim Topoisomerasa, deixant les cadenes accessibles a altres enzims, i es mantenen obertes per efecte de l'enzim Helicasa.

Els nucleòtids s'uneixen els uns als altres mitjançant enllaços entre el grup OH del carboni 3 i el grup fosfat. L'ADN-polimerassa III, l'encarregada de posar i ajuntar els nucleòtids, només pot dur a terme la duplicació si parteix d'un nucleòtid: necessita partir del grup OH del nucleòtid anterior. Però l'ARN-polimerassa sí que pot i, per tant, l'ARN-polimerassa posa uns quants nucleòtids i llavors l'ADN-polimerassa III pot continuar. Després, un altre enzim, l'ADN-polimerassa I, substitueix els fragments d'ARN per fragments d'ADN, i l'ADN-ligasa tanca l'espai entre aquest nou fragment d'ADN i el que ja estava posat.

### Preparació de l'Error-prone-PCR

La PCR és una tècnica utilitzada en la Biologia Molecular que s'utilitza per obtenir moltes còpies d'una seqüència d'ADN d'interès. Com més còpies tenim, més fàcil és treballar amb aquesta seqüència. Podem diferenciar tres grans passos en el procés de la PCR, els quals es repeteixen 30 vegades en una màquina que pot refredar i escalfar els tubs en molt poc temps:

1. **Desnaturalització** a 94 °C: les dues cadenes d'ADN s'obren, quedant separades.
2. **Aparellament** a uns 50°C: s'utilitzen dos encebadors, seqüències d'ADN: un marca el principi del fragment a amplificar i l'altre el final. Aquests encebadors es fixen en una seqüència de l'ADN que li sigui complementària.
3. **Extensió** a 72°C: aquesta és la temperatura ideal per la Taq-polimerassa, una DN-polimerassa més resistent a la calor, fet que és útil perquè ha de suportar els 94°C del primer pas. Aquest enzim es pot adherir al final de l'encebador i pot començar la duplicació de l'ADN.



Imatge 5: Màquina de cicles tèrmics per a PCR  
(Font: Imatge pròpia)

Com es pot veure, en la PCR només s'utilitza un enzim, la Taq-polimerassa, i no més de 12 com en la duplicació de l'ADN. Això és perquè els enzims són molt sensibles a les temperatures, al pH... Així, com menys enzims tinguem més segurs estarem que la PCR

donarà resultat. A més, els enzims són molt cars. Seguidament hi ha una taula on s'exposen els enzims més importants en aquest procés i els canvis que s'han fet per tal de no utilitzar-ne la majoria:

Duplicació de l'ADN	PCR
Topoisomerasa i helicasa	Calor; a 94°C, les cadenes se separen i queden obertes
ADN-polimerassa III	Taq-polimerassa (tipus d'ADN-polimerassa)
ARN-polimerassa	No cal; hi posem encebadors
ADN-polimerassa I	No cal; els encebadors són seqüències d'ADN
ADN-ligasa	No cal; no s'ha hagut de treure una seqüència d'ARN per una d'ADN i, per tant, no s'ha de lligar aquest nou fragment amb la resta de la cadena

Per a aconseguir mutacions vam dur a terme dos canvis en la PCR normal:

1. Vam afegir  $MnCl_2$ : El  $Mg^{+2}$  estabilitza l'estructura dels àcids nucleics i per això s'acostuma a afegir en els tampons de PCR. El  $Mn^{2+}$  no duu a terme aquesta funció correctament i fa que l'ADN-polimerassa no treballi tan bé i, per tant, és més probable que introdueixi errors.
2. Vam canviar la proporció entre purines i pirimidines: enlloc de fer 1:1, vam fer 1 pirimidina (C, T):5 purines (A, G). D'aquesta manera, en un moment determinat l'ADN-polimerassa no trobarà el nucleòtid que necessita i, com que també està afectada pel  $MnCl_2$ , introduirà un nucleòtid equivocat per tal de poder continuar la duplicació.

Després de fer aquests canvis, els resultats esperats eren de 2 a 5 mutacions per gen (resultat mitjà de l'error-prone-PCR), les quals serien totalment aleatòries. En cada cicle es creen noves mutacions, que se sumen a les mutacions anteriors.

### Productes

Els productes que vam utilitzar, i les seves quantitats corresponents, són els següents:

Encebador CyRI	1,75 µL
Encebador CyPstI	1,75 µL
Tampó 10 x PCR	17,5 µL
dNTP error prone	3,5 µL
$MnCl_2$	7,0 µL
Aigua destil·lada	138,25 µL
<b>Total Master Mix</b>	<b>150 µL</b>

Naturalment, en aquests productes hi falta el més important de tots, el patró d'ADN. No l'afegim a la mescla inicial perquè necessitem fer un control de la PCR, el qual veurem en el

gel d'agarosa. Els controls són molt importants perquè són experiments paral·lels al principal on es veu si els productes usats estan bé o no. Si els controls no ens surten com ens esperàvem, no podem donar credibilitat al nostre experiment. Si esperem trobar alguna cosa al control, l'anomenarem "control positiu", i si no esperem trobar-hi res, l'anomenarem "control negatiu".

Per això, aquests 150 µL de la Master Mix els dividirem en tres tubs Eppendorf:

1. 49 µL de la Master Mix + 1 µL d'H<sub>2</sub>O. Aquest és el nostre control negatiu perquè no esperem trobar-hi producte de PCR, ja que no hi ha patró.
2. 49 µL de Master Mix + 1 µL de patró amb el gen HisF
3. 49 µL de Master Mix + 1 µL de patró amb el gen HisF

Els tubs 2 i 3 són el mateix; ho fem per tenir més variabilitat de mutacions.

### ***Programa de la PCR***

- 2 min. a 95°C
- 30 seg. a 95°C
- 30 seg. a 50°C
- 60 seg. a 72°C
- 10 min. a 72°C

Els passos 2, 3 i 4 s'han de repetir 30 cops.

### ***Resultats***

Després de fer la PCR ja tenim el nostre gen hisF modificat, el qual anomenarem hisF\*. El gen més alguns fragments a davant i darrere del gen que ens permeten assegurar que el gen està sencer en el fragment fan un total de 1000 parells de bases.

## **Gel d'agarosa. Electroforesi**

L'electroforesi és una tècnica utilitzada en Biologia Molecular que serveix per separar diferents fragments d'ADN o diferents proteïnes segons la seva llargada i utilitzant que l'ADN i les proteïnes tenen càrrega elèctrica. El gel d'electroforesi és una dissolució d'agarosa en un tampó. Després d'escalfar aquesta dissolució i deixar-la reposar, obtenim un producte semblant a una gelatina que forma una xarxa.

La preparació d'electroforesi té un pol positiu i un pol negatiu. Com que treballem en ADN, em centraré en aquest cas. L'ADN té càrrega negativa i per això el dipositem dins uns pous que fem al gel d'electroforesi i que es troben al costat del pol negatiu. D'aquesta manera, quan s'aplica un corrent entre els pols, l'ADN es desplaça cap al pol positiu, havent de travessar la xarxa d'agarosa. Els fragments més petits podran travessar més fàcilment el gel i per això es trobaran més lluny del punt d'origen. Així podem separar fragments d'ADN segons la seva mida.

Nosaltres la vam utilitzar en aquest cas per veure si podíem identificar un producte de PCR. Com que només volem veure el producte d'un altre procés, aquest tipus de gel s'anomena "gel d'agarosa analític". Després de la PCR hauríem de ser capaços d'identificar una banda a l'altura dels 1000 pb, ja que hem amplificat tant aquest fragment que el conjunt de tots els fragments és visible a ull humà.

### Preparació del gel

1. Nosaltres volem un gel d'Agarosa a l'1%; com més gran és el tant per cent més dens és el gel i més els costa als fragments d'ADN avançar pel gel. Dissoldrem 1 gram d'Agarosa en 100 ml del tampó Tris-Borat-EDTA (TEB). Ho hem d'escalfar i barrejar fins que es dissolgui tota l'Agarosa.
2. Per preparar l'aparell d'electroforesi posem la pinta que farà els diferents pous al costat del pol negatiu i limitem l'espai del gel de l'electroforesi, de manera que tingui forma rectangular.
3. Quan la dissolució del microones arribi a uns 60°C, hi afegirem 5 µL a 10mg/ml d'un producte cancerigen anomenat Ethidiumbromide, que ens permetrà veure totes les bandes d'ADN sota llum ultraviolada, ja que tenyeix l'ADN de color taronja gràcies a la seva propietat de fixar-se entre les diferents bases de l'ADN. S'ha d'anar amb molt de compte a l'hora d'usar-lo perquè si ens toca la pell també pot entrar a les nostres cèl·lules i fixar-se al nostre ADN, augmentant el risc de patir mutacions durant la duplicació.
4. Ja es pot abocar tota la dissolució al seu recipient dins l'aparell d'electroforesi. La dissolució no podrà ocupar el lloc on es troben les puntes de la pinta i així obtindrem els pous. S'ha de deixar reposar uns 30-45 min.

### Preparació del contingut dels pous

1. Agafarem 5 µL de cada tub Eppendorf després d'haver fet la PCR. A cada un hi afegirem 5 µL d'un tampó blau de càrrega (LD1), el qual conté un pigment blau (ens servirà per saber com està corrent el gel) i glicerol (farà que la mostra sigui densa i no surti del pou).
2. Quan el gel estigui solidificat traiem la pinta i les estructures que en limitaven l'espai.
3. Aboquem dins l'aparell d'electroforesi dissolució TBE, la qual ha d'arribar a cobrir el gel d'electroforesi. L'utilitzem com a dissolució conductora de l'electricitat.
4. Aboquem les nostres mostres als pous utilitzant micropipetes. El primer pou el reservem pel marcador, el qual és una preparació amb diferents mides d'ADN conegudes dissoltes. Agafant com a comparació el marcador, podem saber quina llargada tenen les nostres mostres d'ADN.
5. L'aparell d'electroforesi s'ha de connectar a una màquina de corrent elèctric a uns 130V.

### Resultats

Després d'esperar que les mostres correguessin pel gel, vam analitzar el nostre gel sota llum ultraviolada.

Per sorpresa nostra, al gel només hi vam poder detectar el marcador. Així que vam tornar a preparar el gel i també les mostres, i no vam veure res altre cop. Què havia passat?

El que vam fer després va ser fer tota una sèrie de controls amb tots els productes que havíem utilitzat quan fèiem la PCR. Vam veure que només podíem veure bandes d'ADN al gel d'electroforesi quan no posàvem el  $MnCl_2$ . Per això vam endevinar que era un problema de la concentració de  $MnCl_2$ . Una alta concentració d'aquest producte no permet treballar a l'ADN-polimerassa i per això no teníem producte de PCR.

Ho vam tornar a fer amb una altra concentració de  $MnCl_2$  i vam obtenir una clara marca als 1000 pb. La PCR havia funcionat i, per tant, ja teníem el nostre gen hisF\*.



## Aïllament de plasmidis

Ara ja tenim el gen. El que volem ara és un plasmidi on introduir el nostre gen. Per això el següent és dur a terme una aïllament de plasmidis d'una colònia de bacteris E. Coli. El plasmidi d'aquests bacteris té un gen de resistència a l'ampicil·lina que ens servirà al final per saber del cert si els bacteris tenen el plasmidi o no.

### Tècniques en el procés d'aïllament de plasmidis

Dels bacteris amb els que ara treballem només en necessitem el plasmidi. Per tant, ens hem de desfer del cromosoma principal dels bacteris. Per fer-ho utilitzarem canvis en el pH.

L'ADN dels bacteris es desnaturalitza a 12-12.5. Això vol dir que desapareix l'estructura secundària, terciària i quarta, de manera que ja no pot dur a terme la seva funció. L'única que es manté és l'estructura primària. A més, aquest pH tan alt més NaOH i SDS (Sodium Dodecyl Sulfat) destrueixen la membrana de les cèl·lules.

Si el tornem a naturalitzar, només el plasmidi pot tornar a la seva forma original, mentre que el cromosoma es desintegra a causa de la seva grandària, i les proteïnes també, i acaben precipitant. La centrifugació separa aquestes substàncies degradades i altres parts de la cèl·lula, quedant el plasmidi com a sobrenedant, dissolt amb altres substàncies.

Com que més endavant farem servir enzims i aquests necessiten unes condicions molt específiques per poder funcionar, separarem el plasmidi de les salts dissoltes (com el NaOH i el SDS) amb les columnes amb fibra de vidre. Aquestes columnes retenen els àcids nucleics i deixen passar la resta. Si hi afegim un tampó amb poca concentració salina podrem desenganxar l'ADN de la fibra de vidre.

### Procés de l'aïllament de plasmidis

- Centrifuguem 1.5 ml de cèl·lules de bacteris a la màxima potència (13.400 rpm) durant un minut. Com que les cèl·lules són el sediment, podem llençar el sobrenedant.
- Resuspenem els bacteris en 250 µL del Tampó 1, el qual ha de fer augmentar el pH, utilitzant el vortex durant 3 minuts.
- Afegim 250 µL del Tampó 2, el qual conté el NaOH i el SDS, i invertim el tub de 4 a 6 cops i esperem 4 minuts. No podem utilitzar el vortex perquè fariem malbé l'ADN.
- Afegim 250 µL del Tampó 3, el qual redueix el pH i invertim el tub immediatament de 4 a 6 cops. Com que comencen a precipitar l'ADN, les proteïnes i altres parts de la cèl·lula, anem movent la solució per evitar precipitacions localitzades.
- Ho centrifuguem 10 min. a 13.000 rpm. Com a resultat tindrem un sediment blanc. Els plasmidis estaran al sobrenedant.
- Aboquem el sobrenedant a una columna, i aquesta a un tub per posar-ho a la centrifugadora durant 60 s a màxima velocitat. Tot el que no sigui ADN travessarà la columna i caurà al tub i, per tant, ho podem llançar. Tornem a posar la columna dins del tub.
- Afegim 0.70 ml del Tampó 4, el qual netejarà la columna de possibles altres substàncies que no siguin ADN. Ho centrifuguem durant 60 s a màxima velocitat. Tornem a descartar el líquid del tub.
- Centrifuguem la columna altra vegada perquè els residus del Tampó 4 només poden marxar totalment un cop ens desfem del filtrat. És important que no quedin residus

perquè entre aquests es troba l'etanol, el qual podria aturar futures reaccions enzimàtiques.

- Posem la columna en un tub Eppendorf. Afegim 100 µL del Tampó 5, el qual té poca concentració salina. Fins i tot podríem afegir aigua. Ho deixem reposar 1 min i centrifugar 1 min.

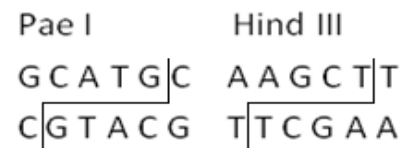
En el tub Eppendorf hi tenim una solució aquosa que conté els nostres plasmidis. Ja podem llençar la columna.

## Restricció

Un cop tenim els fragments amb els nostres fragments amb el gen *hisF\** i els plasmidis a punt, el següent pas és inserir el gen al plasmidi. Per poder-ho dur a terme realitzarem dos processos: la restricció i la lligació.

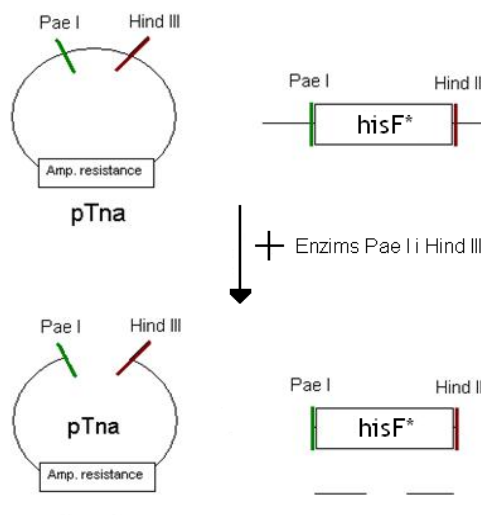
El que volem aconseguir amb la restricció és aïllar el nostre gen del fragment de la PCR i també aconseguir tallar el plasmidi per tal d'inserir-hi, en un futur, el nostre gen aïllat. Per dur a terme aquest procés necessitem els anomenats enzims de restricció, on cada un d'aquests enzims reconeix una seqüència específica d'ADN. L'enzim s'enganxa en la seva seqüència i fa un tall en les dues cadenes, procés que també podem anomenar digestió. Sempre talla la mateixa seqüència i de la mateixa manera.

Nosaltres vam utilitzar dos enzims diferents, el Pae I i el Hind III. Com es pot veure a l'esquema, com que l'enzim Pae I reconeix GCATGC, s'enganxa a la molècula d'ADN i hi fa un tall seguint la línia representada a l'esquema. El mateix fa l'enzim Hind III amb la seva seqüència. Aquests finals s'anomenen "sticky ends".



Imatge 6: Seqüència específica de restricció de Pae I i Hind III amb "sticky ends" (Font: Dibuix realitzat amb Paint)

Aquestes dues seqüències es poden trobar tant als fragments de la PCR com al plasmidi, tal com es marca a la Imatge 7, en verd per l'enzim Pae I i en vermell per l'enzim Hind III.



Imatge 7: Fragments resultants de la restricció del plasmidi i del producte de la PCR (Font: Dibuix realitzat amb Paint)

### Procés per a la restricció

Utilitzem dos tubs Eppendorf, un amb 25 µL del producte de la PCR i l'altre amb 25 µL dels plasmidis. En tots dos hi afegirem 3 µL de Tampó Tango, 1 µL de l'enzim Pae I i 1 µL de l'enzim Hind III.

Fem un spin a la centrifugadora durant uns segons i ho deixem digerir durant una hora al Thermo-Block a 37°C, la temperatura que necessiten els enzims per actuar.

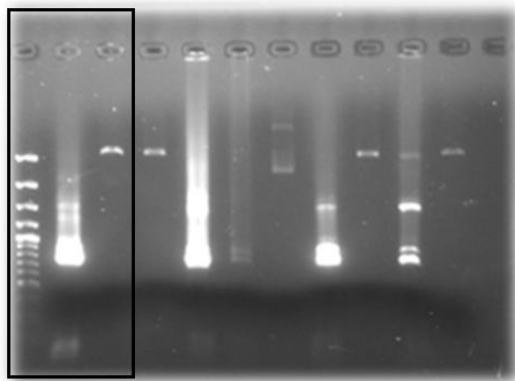
En la Imatge 7 podíem veure el resultat de la digestió. El nostre gen hisF\* també es pot anomenar insert, i el nostre plasmidi també es pot dir vector.

### Aïllament de fragments d'interès a partir d'un gel d'agarosa

A diferència del primer gel que hem fet, que era d'anàlisi, el que ara farem és un “gel d'agarosa preparatiu”, que vol dir que no el llencem després de comprovar els nostres resultats, sinó que agafem certes mostres de fragments del gel. És per això que a l'hora d'omplir els pous no posem mostres simplement representatives sinó que hi posem tota la mostra. Tot i així, la preparació del gel és exactament igual que l'anterior.

En la Imatge 8 podem veure el resultat del nostre gel. Els fragments d'ADN estan enfocats amb llum ultraviolada i els podem veure clarament gràcies a l'Ethidiumbromide. La primera columna és el marcador, i les altres deu columnes representen cinc vegades el mateix experiment. Dues d'aquestes juntament amb el marcador estan dins del requadre negre per tal de centrar l'atenció en aquestes columnes i no en les altres, les quals són el mateix però repetit per diferents grups.

Mitjançant el marcador sabem que els fragments de la tercera columna tenen una llargada de 3000 pb (són els nostres vectors), i que els fragments de la segona tenen 800 pb (el nostre insert). A la mateixa columna de l'insert hi podem trobar altres fragments en franges més superiors; això és perquè hem afegit molta quantitat d'ADN als pous i no tota ha pogut avançar de la mateixa manera. També hi podem veure una banda en la part inferior; aquests fragments tenen una llargada d'uns 100 pb i són els fragments del costat del gen hisF\*, tal com es podia veure en la Imatge 7. Com a resultat de la restricció del plasmidi havíem d'obtenir el vector i un fragment més petit de 20 pb, el qual no podem veure perquè, en ser tan petit, ha sortit del gel. Haguéssim pogut esperar menys i així haguéssim pogut veure aquests fragments petits, però el que més ens interessava era tenir els fragments de diferents mides ben separats, i per això ho deixem córrer més temps.



Imatge 8: Gel d'electroforesi amb els productes de la restricció (Font: Imatge pròpia)

De tots aquests fragments els únics que ens interessin són el gen *hisF\** i el plasmidi. Així que el que fem és aïllar-los del gel d'agarosa. Com que a l'hora de fer l'extracció treballarem amb rajos ultraviolats i amb gel que conté Ethidiumbromide, haurem de portar una màscara que ens cobreixi la cara i també guants especials per les mans. Els passos són aquests:

1. Extraïem el tros del gel d'agarosa que conté els gens *hisF\** i els plasmidis amb un bisturí net i afilat, i posem cada mostra a un tub Eppendorf. Hem d'intentar que la mostra tingui com menys gel d'agarosa possible, ja que després l'haurem de dissoldre i serà més ràpid si hi ha menys agarosa.
2. Pesem el gel. Per cada 100 mg de gel, hem d'afegir 300 µL del Tampó QG, el qual dissoldrà l'agarosa de la mostra. Tant la mostra amb els nostres inserts com amb les plasmidis pesaven 0,21 g. Així, havíem d'afegir 630 µL del Tampó QG.
3. Per ajudar a la dissolució del gel, podem utilitzar el vòrtex. És important que no quedi cap mostra d'agarosa.

El procés que comença ara és similar al que vam dur a terme per aïllar els plasmidis:

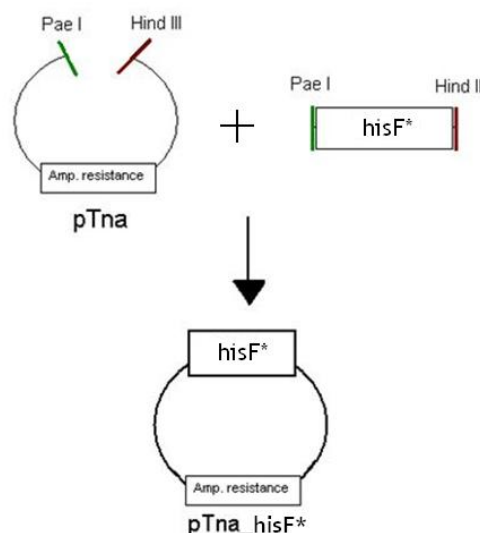
4. Posem el contingut de l'Eppendorf a una columna QIAquick, i aquesta a un tub de 2 ml. Així, tindrem dues preparacions: una pel gen *hisF\** i l'altre pels plasmidis. Posem aquests dos tubs a centrifugar durant un minut.
5. Llancem el líquid que hagi travessat la columna i tornem a col·locar-la al mateix tub.
6. Afegim al tub 0,70 ml de Tampó PE a la columna, el qual la netejarà, i centrifuguem 1 min.
7. Descartem el líquid filtrat i centrifuguem altre cop la preparació durant un altre minut a 13.000 rpm.
8. Posem la columna neta a un tub de 1,5 ml.
9. Per desenganxar l'ADN de la columna, afegim 50 µL del Tampó EB o aigua i ho centrifuguem durant un minut.

Per veure si l'aïllament s'ha realitzat correctament, duem a terme un altre gel d'agarosa, on només han d'aparèixer una banda a 3000 pb i a 800 pb. El gel és també de l'1%.

## Lligació

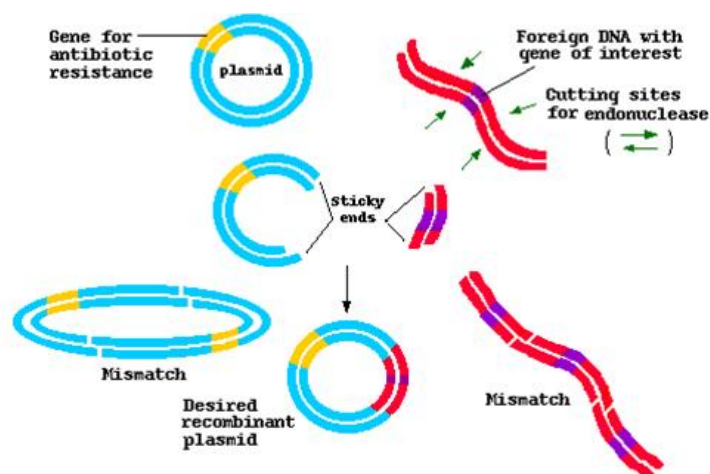
Tenim el nostre gen, tenim el plasmidi a punt. El que només falta és ajuntar-los. Per a això, utilitzarem la lligació, també anomenada clonació.

Com que vam tallar el producte de la PCR i el plasmidi amb els mateixos enzims, ara els podem fer encaixar, ja que els "sticky ends" són complementaris. Llavors l'enzim lligasa connectarà els finals oberts i tancarà els espais entre els dos fragments. El cas representat a la dreta és el que desitgem, però també hi haurà casos en què, com veiem a la Imatge 10, alguns vectors s'ajuntaran entre ells o alguns inserts s'ajuntaran entre ells. Aquests dos últims casos no ens interessin, però no podem fer res per evitar-los. Tot i així, un bacteri amb el cas de múltiples vectors sense l'insert no podrà codificar per una proteïna amb funció de HisA i un bacteri amb múltiples inserts però sense gen per a la



Imatge 9: Plasmidi desitjat resultant de la lligació del vector i l'insert (Font: Dibuix realitzat amb Paint)

resistència a l'Ampicil·lina no podrà viure en un medi amb Ampicil·lina, fet que comporta que aquests dos suposats bacteris no puguin créixer en la plata del nostre experiment que prepararem més endavant. D'aquesta manera, només sobreviuran els bacteris amb el plasmidi correcte.



Imatge 10: Representació dels possibles productes de la lligació  
(Font: <<http://www.bio.davidson.edu>>)

### Preparació

Els productes utilitzats, amb les seves respectives quantitats, són els següents:

Tampó pel bon funcionament de la lligasa	2 µL
Plasmidi després de la digestió	1 µL
Gen hisF*	5 µL
Enzim lligasa	2 µL

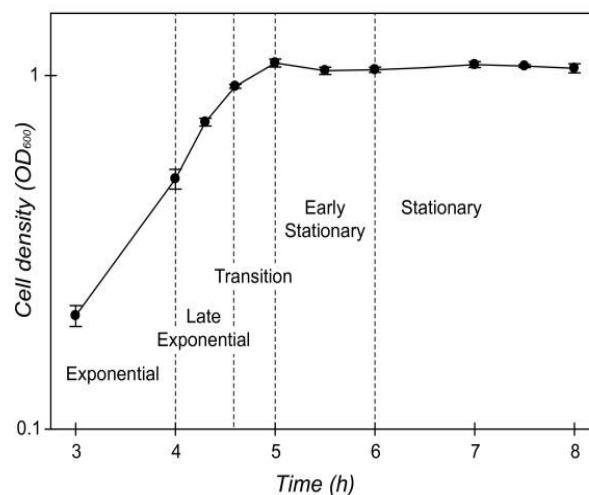
Centrifuguem la mescla durant uns segons i la deixem a 37 °C durant 1 hora. Un cop aquest temps tindrem el nostre plasmidi amb el gen hisF\*, el qual podem anomenar pTna\_hisF\*.

### **Cèl·lules químicament competents**

El que ara necessitem són bacteris sense el gen de la HisA per a introduir-hi el nostre plasmidi i dur a terme el nostre experiment: veure si sobreviuen en un medi sense Histidina.

Per a això començarem a fer créixer bacteris: mesurarem 50 ml d'LB (una solució amb tot el que els bacteris necessiten per a créixer) i els abocarem a un flascó Erlenmeyer de 200 ml conjuntament amb algunes cèl·lules. Hem de deixar créixer les cèl·lules durant la nit a una temperatura de 37°C movent-se a 200 rpm.

Un cop tenim un gran nombre de cèl·lules, les hem de fer químicament competents, que vol dir que seran capaces d'incorporar ADN provinent de l'exterior. Per aconseguir-ho, duem a terme aquests passos:



Imatge 11: Representació de l'augment de l'OD<sub>600</sub> al llarg del temps i fases del creixement (Font: <<http://www.bio.davidson.edu>>)

1. Canviem de medi als bacteris: els passem a 100 ml de TY-Media, el qual és igual que l'LB però conté més hidrocarbons, component que propicia el creixement més accelerat dels bacteris. Els hem de deixar reproduir-se a 37°C. Per saber fins quan els hem de deixar créixer, primer hem de conèixer el creixement dels bacteris.

En el gràfic de sota (Imatge 11) hi és representat la densitat òptica (OD en anglès) del medi amb bacteris respecte el temps. L'OD es calcula enviant una ona longitud d'ona concreta (la més comuna és la de 600 nm) a través del medi de bacteris. Al medi hi entra una llum del 100% i en surt un tant per cent de x. Com més gran és x, menys bacteris tenim al medi, ja que la llum ha passat sense massa problemes. Amb aquest valor es pot trobar l'OD. Si l'OD<sub>600</sub> és igual a 1, llavors tenim 10<sup>9</sup> cèl·lules/ml.

En el gràfic veiem que en la primera fase, anomenada fase exponencial, el nombre de bacteris creix molt en poques hores. Arriba un punt, però, on els nutrients comencen a mancar, i és llavors quan ens trobem a la fase estacionària. Si deixéssim passar més temps, els bacteris es començarien a morir.

Nosaltres volem que els nostres bacteris es trobin a la fase exponencial, ja que es troben en millor estat per poder ser convertits en cèl·lules químicament competents, especialment de 0,6 a 0,8 d'OD<sub>600</sub>.

Temps	Inicial	20 min	40 min	50 min
OD <sub>600</sub>	0,179	0,372	0,53	0,7

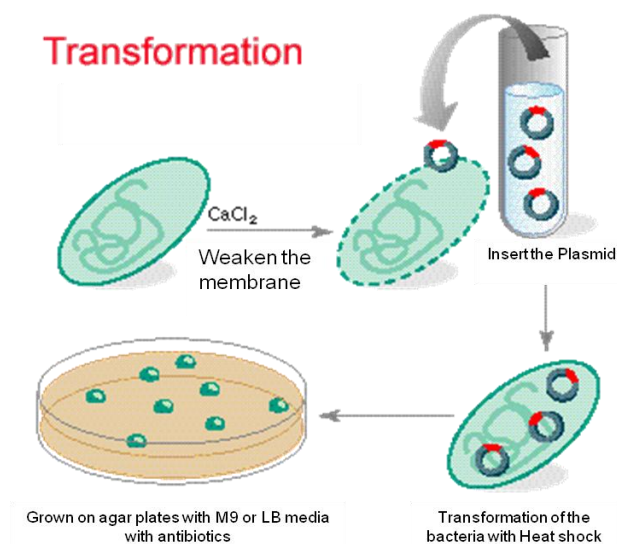
Tal com ens indica l'OD, al cap de 50 min les cèl·lules estan en un bon punt per continuar el procés.

2. Afegim 2 ml de MnCl<sub>2</sub> (1 M) i ho barregem durant 10 min a 37°C. El que aconseguim amb aquest producte és que la membrana es comenci a afeblir.
3. Deixem refredar les cèl·lules 1 h en gel.
4. Fem un spin durant 5 min a 5000 rpm i a 4°C i llancem el sobrenedant, ja que les nostres cèl·lules es troben al sediment.

5. Hi afegim 35 ml d'AC, el qual conté  $\text{CaCl}_2$ , substància que també debilita la membrana, i resuspenem el sediment per tal que es barregi amb l'AC. Ho fem amb una pipeta esterilitzada d'1 ml, fent entrar i sortir el líquid de la pipeta, per tal d'aixecar les cèl·lules. No utilitzem el vortex perquè faríem malbé les cèl·lules.
6. Ho deixem reposar en gel per una altra hora.
7. Fem un spin durant 5 min a 5000 rpm i a  $4^\circ\text{C}$  i llancem el sobrenedant.
8. Hi afegim 4 ml d'ACG, el qual conté  $\text{CaCl}_2$  i glicerol, substància que també debilita la membrana, i resuspenem el sediment.

## Transformació de les cèl·lules químicament competents

Una cèl·lula transformada significa que incorpora i expressa material genètic provinent de l'exterior. Ara ja tenim les parets de les cèl·lules debilitades i amb forats i, per tant, serà més senzill fer-hi entrar el nostre plasmidi.



Imatge 12: Transformació de cèl·lules químicament competents (Font: <<http://www.biomedcentral.com>>)

Així, tindrem un cultiu de bacteris amb el nostre plasmidi. A més, per tal de tenir controls, introduïrem un plasmidi amb el gen *hisA* (pTna\_ hisA) a un altre cultiu de bacteris (control positiu) i introduïrem un plasmidi amb el gen *hisF* (pTna\_ hisF) a un altre cultiu de bacteris (control negatiu). Aquests dos plasmidis també contenen el gen de la resistència a l'Ampicil·lina. En tots tres casos, el procediment serà el mateix; utilitzarem el Heat Shock, canvis bruscos de temperatura, cosa que permet que el plasmidi entri dins dels forats de la paret cel·lular:

1. Descongelem les nostres preparacions amb els bacteris, tres mostres de 200  $\mu\text{L}$  cada un.
2. Afegim 1  $\mu\text{L}$  del plasmidi (en cada preparació, un dels tres anomenats anteriorment) i ho mantenim en gel durant 5 min.
3. Ho escalfem a  $42^\circ\text{C}$  durant 42 s i ho refredem en gel durant 5 min. El plasmidi ja està a l'interior de la cèl·lula.
4. Hi afegim 800  $\mu\text{L}$  del medi LB i ho deixem incubar durant 45 min a  $37^\circ\text{C}$  mentre es va sacsejant. Volem que les cèl·lules creixin.

## Preparació de les Plaques de Petri amb els diferents medis de cultiu

Per poder dur a terme el nostre experiment i els seus respectius controls necessitarem dos medis: un medi complet (medi LB) i un medi mínim (M9), tots dos amb Ampicil·lina. Cada un tindrà un sistema de preparació diferent.

### Medi LB

El medi LB està pensat per E. coli i conté:

- Tryptone: proteïnes i aminoàcids (ja té, per tant, Histidina)
- Extracte de llevat: font de proteïnes addicionals, vitamines, minerals i sucres
- NaCl
- Agar: solidifica la solució

L'únic que hem de fer és abocar en una proveta 6g d'un producte ja preparat que conté aquests components, i acabar d'omplir la proveta amb H<sub>2</sub>O fins que arribem als 125 ml. Necessitem 5 plaques de medi LB: un serà amb tant sols LB i els altres quatre amb LB + Ampicil·lina. Així, farem una Autoclau (escalfar-ho a 121°C durant 20 min i a 2 bars), ho deixem refredar fins a 60°C, i aboquem part de la dissolució a una placa de Petri. Llavors afegim 80 µL d'Amp a la resta de la dissolució i la repartim en 4 plaques. Deixem refredar les 5 plaques.

### Medi mínim

Hi ha diferents tipus de medis mínims; el nostre és el 9. Aquest conté:

- Glucosa
- Molta quantitat de sals (Ca<sup>2+</sup>, Mg<sup>2+</sup>, N<sup>3</sup>,...) per tal que es puguin construir els aminoàcids i les proteïnes, ja que no n'hi ha al medi.

Barregem 130,5 ml d'H<sub>2</sub>O i 2,25 g d'Agar, fem una Autoclau i ho deixem refredar fins a 60°C. Llavors, mentre la dissolució va movent-se, hi afegim:

M9 Salt 10X (conté les sals)	15 mL
MgSO <sub>4</sub>	150 µL
Glucosa	1,5 mL
Thiamine	150 µL
CaCl <sub>2</sub>	1,5 mL
Ampicil·lina	150 µL

L'únic que ens queda fer és repartir la dissolució en cinc plaques de Petri i deixar-les refredar.



### Extensió dels bacteris en els diferents medis

	Cèl·lules químicament competents (control -)	$\Delta$ HisA + ptna_hisA (control +)	$\Delta$ HisA + ptna_hisF (control -)	$\Delta$ HisA + ptna_hisF* (experiment)
LB	(LB sense Amp) 50 $\mu$ L	50 $\mu$ L	50 $\mu$ L	100 $\mu$ L
	50 $\mu$ L			
M9		50 $\mu$ L	50 $\mu$ L	200 $\mu$ L
				200 $\mu$ L
				200 $\mu$ L

Aquest quadre representa la quantitat que posem de cada tipus de bacteri en els diferents medis.

### Utilitat de cada placa

1. Els medis amb les cèl·lules químicament competents els utilitzem per veure que les cèl·lules en el medi amb Amp moren, ja que no tenen el plasmidi amb el gen de la resistència d'aquest antibiòtic. És per això que només el realitzem en medi LB, ja que la causa de la seva mort en el medi M9 es deuria també a la falta del gen hisA.
2. Els bacteris  $\Delta$ HisA + ptna\_hisA els utilitzem com a control positiu perquè esperem que en el medi M9 sobrevisquin, ja que tenen el gen per sintetitzar-se la proteïna que els permetrà obtenir Histidina.
3. Els bacteris  $\Delta$ HisA + ptna\_hisF els utilitzem com a control negatiu perquè esperem que en el medi M9 no puguin créixer a falta del gen per sintetitzar-se la proteïna que els permetria obtenir Histidina.
4. Finalment, els bacteris  $\Delta$ HisA + ptna\_hisF\* en el M9 són el nostre veritable experiment i per això en tenim 3 plaques i amb més quantitat de bacteris.
5. En el medi LB tots els bacteris haurien de créixer i, per tant, ens serveix com a control positiu. Posem Amp als medis perquè així qualsevol bacteri que no s'hagi transformat correctament es mori.

## Resultats

### Resultats esperats

D'aquesta manera, els nostres resultats esperats eren aquests:

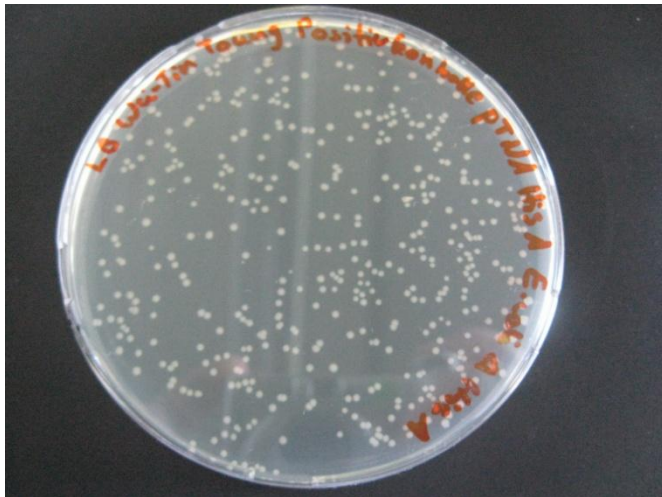
	Cèl·lules químicament competents (control -)	$\Delta$ HisA + ptna_hisA (control +)	$\Delta$ HisA + ptna_hisF (control -)	$\Delta$ HisA + ptna_hisF* (experiment)
LB	(LB sense Amp) +	+	+	+
	-			
M9		+	-	? (+)
				? (+)
				? (+)

El signe + representa que hi esperem creixement i el signe - que no hi esperem cap bacteri.

Els quadres amb els ? són el nostre experiment, del qual no sabem el resultat. Tot i així, l'experiment tindrà els resultats desitjats si hi ha creixement, d'aquí que al costat de l'interrogant hi hagi un (+).

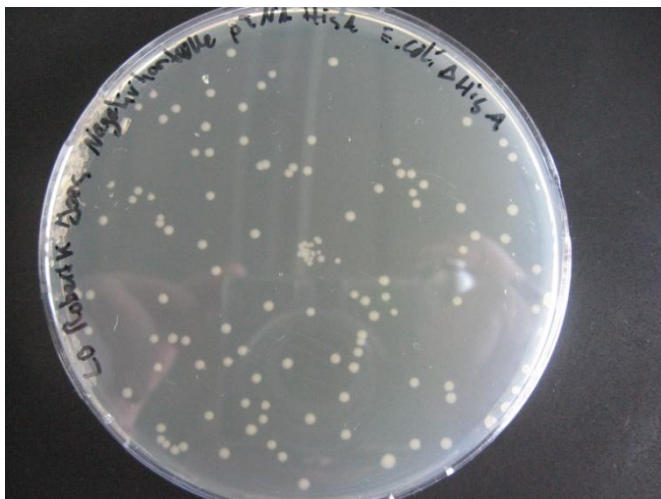
## Resultats obtinguts

Després de créixer durant tota la nit a 37°C, vam poder observar aquests resultats:



Imatge 12: LB, E.coli  $\Delta$  HisA + pTna\_hisA

Com podem veure en aquesta imatge, hi ha hagut molt creixement en el medi LB amb els bacteris que tenien el plasmidi amb el gen hisA, cosa que ja estava prevista.



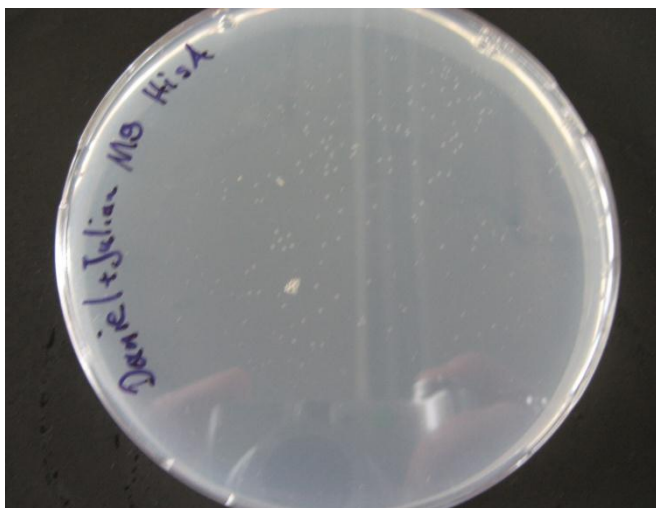
Imatge 13: LB, E.coli  $\Delta$  HisA + pTna\_hisF

Tot i ser en un medi LB com en la imatge anterior, aquí podem observar que no hi ha hagut tant creixement. Això és causat pel tipus de promotor dels gens hisA, hisF i hisF\*: el promotor d'aquests gens fa que aquests es transcriguin en tot moment. Així, tot i tenir Histidina al medi, els bacteris del primer cultiu en continuen produint i, com que en tenen més quantitat, poden créixer més.



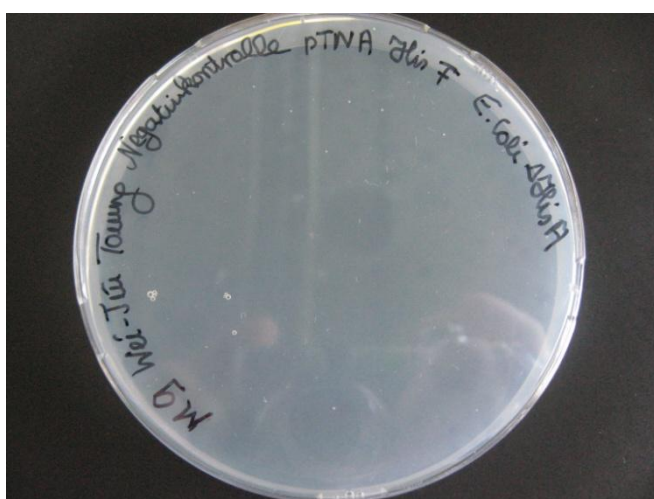
Imatge 14: LB, E.coli  $\Delta$  HisA + pTna\_hisF\*

Aquest és el mateix cas que en la imatge 13: hi ha creixement però no tant com en el primer cas.



Imatge 15: M9, E.coli  $\Delta$  HisA + pTna\_hisA

Podem veure com els bacteris amb el gen *hisA* han sobreviscut en un medi sense Histidina, tal i com estava previst. Tot i així, veiem que el seu creixement és molt inferior al corresponent en el medi LB. Aquest fet és fàcilment explicable: ara el bacteri s'ha de sintetitzar totes les substàncies que necessita, fet que comporta que creixi més lentament.



Imatge 16: M9, E.coli  $\Delta$  HisA + pTna\_hisF

En aquesta placa no hi podem veure creixement. Ens podria semblar que n'hi ha, però vist de prop es veu que són bombolles. Tal i com estava previst, si aquests bacteris no es poden sintetitzar la *HisA* en un medi que no els proporciona Histidina, es moriran, ja que no podran acabar produint aquest aminoàcid bàsic per a la creació de proteïnes.



Imatge 17: M9, E.coli  $\Delta$  HisA + pTna\_hisF\*

Aquesta és una de les plaques del nostre experiment amb el plasmidi que vam crear pTna\_hisF\*. Pel que hi podem veure, hi ha creixement, tot i que no molt abundant. Això no és estrany perquè estem en un medi M9. Si ho comparem, hi ha més creixement en la Imatge 15 que aquí. Aquest fet segurament es deu que la proteïna HisF\* que hem aconseguit que tingui la funció de la *HisA*, no és tan efectiva com la *HisA*.

## Discussió

Com hem pogut veure en aquest últim subapartat, els resultats esperats i els obtinguts coincidien. Això ens indica que tots els controls són correctes i que, per tant, podem confiar amb els resultats del nostre experiment real.

Pel que fa a l'experiment, del qual no podíem endevinar el resultat, hem vist que ha acabat donant que un conjunt de bacteris amb el plasmidi pTna\_hisF\* ha sobreviscut i crescut en el medi mínim, l'M9. Això vol dir que el gen que es troba en aquest plasmidi pot sintetitzar una proteïna que té la capacitat de dur a terme la funció de la HisA. No necessàriament la proteïna HisF\* serà igual estructuralment parlant que la HisA, però almenys en pot fer la funció que, com ja hem vist, no és tan bona com la de la proteïna original, ja que els bacteris amb el pTna\_hisF\* han crescut menys.

Tot i així, cal dir que érem cinc grups que vam realitzar aquest experiment, on cada grup tenia tres plaques amb E.coli  $\Delta$  HisA + pTna\_hisF\* en el medi M9, i només en una placa vam poder veure-hi creixement. Com que en l'error-prone-PCR les mutacions que se'n derivaven eren a l'atzar, les possibilitats que aquestes permetessin que el gen resultant pogués donar lloc a una proteïna amb la funció de la HisA eren també a l'atzar. De fet, hi havia més possibilitats que la proteïna resultant no pogués dur a terme la funció desitjada que no pas que la pogués fer.

Per assegurar la fiabilitat d'aquest experiment, l'hauríem de reproduir des del principi diversos cops. Nosaltres només vam tenir quatre dies per a realitzar aquest experiment, així que no el vam poder repetir.

# Evolució i comparació de l'hemoglobina en diferents espècies

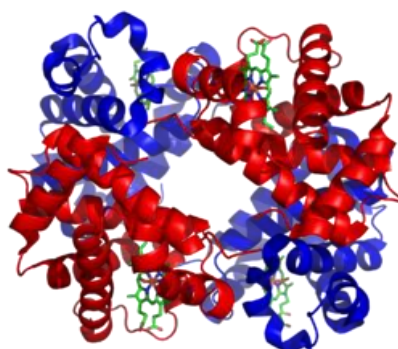
## Introducció

L'objectiu d'aquesta pràctica és comparar les diferents estructures de l'hemoglobina en diferents espècies per tal d'explicar com afecten els mecanismes de l'evolució a les proteïnes. Hem utilitzat l'hemoglobina a l'hora de fer la pràctica perquè és una proteïna altament coneguda, amb una funció important i present en molts animals.

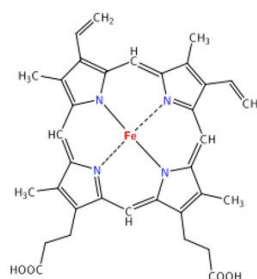
## L'hemoglobina

L'hemoglobina es troba en els glòbuls vermells dels organismes. Cada glòbul vermell conté un nombre aproximat de 280 milions d'hemoglobines. La funció d'aquesta proteïna és transportar l'oxigen dels pulmons als teixits i, següent, transportar el diòxid de carboni dels teixits als pulmons. En el cas dels peixos, l'equivalent dels pulmons són les brànquies.

L'hemoglobina està formada per quatre cadenes polipeptídiques en general, dues  $\alpha$  i dues  $\beta$  (Imatge 1). Les cadenes  $\alpha$  estan formades per 141 aminoàcids, mentre que les  $\beta$  per 146. Les dues classes de cadenes comparteixen estructures secundàries i terciàries similars, cada una formada per vuit hèlix  $\alpha$ . Cada cadena conté, a més a més, un grup hemo (Imatge 2), el qual dóna el color vermell característic de la sang. És la part més important de l'hemoglobina perquè és un component orgànic amb un ferro al centre en el qual s'uneix l'oxigen. Anomenem desoxihemoglobina a l'hemoglobina sense l'oxigen unit al ferro, i oxihemoglobina a la que té l'oxigen unit al ferro.



Imatge 1: Estructura quaternària de l'hemoglobina. Les cadenes  $\alpha$  de vermell i les  $\beta$  de blau. Els grups hemo de verd (Font: <www.molecularstation.com>)



Imatge 2: Grup hemo amb el ferro al centre (Font: <www.molecularstation.com>)

Per tal de facilitar la tasca d'estudi i de comparació de les diferents hemoglobines, només vam agafar una cadena  $\alpha$ , la A, ja que les diferents cadenes de les proteïnes reben lletres diferents seguint l'ordre alfabètic. És important destacar que durant aquest apartat potser parlare d'"hemoglobina" quan en realitat només es tractarà de la cadena A.

## Espècies estudiades

La informació sobre les estructures de les diferents hemoglobines la vam obtenir d'un banc de dades de proteïnes, el PDB (*Protein Data Bank*, en anglès). Allà vam buscar de quines espècies es tenia informació registrada sobre la seva hemoglobina. Per tal d'obtenir uns resultats més interessants en la comparació d'aquesta proteïna en diferents espècies, vam triar animals que representessin, en global, la majoria de grups de vertebrats, ja que després podríem intentar endevinar la relació evolutiva que els uneix.

Així, les espècies estudiades van ser les següents: *Homo sapiens* (humà), *Canis lupus familiaris* (gos), *Anas platyrhynchos* (ànec de collverd), *Columba livia* (colom) i *Thunnus thynnus* (tonyina). L'hemoglobina d'humà i la de gos representen els mamífers, la d'ànec de collverd i la de colom representen els ocells i la de tonyina representa els peixos. No vam poder trobar cap hemoglobina d'amfibi ni de rèptil.

## Apartats de l'estudi de les hemoglobines

Un cop amb les estructures de cada proteïna, ens vam proposar realitzar aquests quatre punts:

1. Visualitzar les diferents hemoglobines amb el programa VMD.
2. Comparar l'estructura primària de les diferents hemoglobines amb el programa ClustalW2 i determinar quina relació evolutiva existeix entre elles.
3. Estudiar la zona menys conservades i les zones perfectament conservades de les proteïnes un cop comparades les estructures primàries.
4. Utilitzar el programa Flex Serv per a obtenir informació sobre la flexibilitat de les proteïnes.



# Visualització de les diferents hemoglobines

## Material utilitzat

Per tal de poder visualitzar una proteïna necessitem un programa i un arxiu amb la informació de l'estructura de la proteïna. El programa que hem utilitzat és el VMD, el qual es pot descarregar gratuïtament a <[www.ks.uiuc.edu/Research/vmd](http://www.ks.uiuc.edu/Research/vmd)>. Aquest és un programa de visualització molecular per visualitzar, modelar, animar i analitzar grans sistemes biomoleculars com proteïnes, àcids nucleics, lípids, etc... utilitzant imatges en 3D. Té una gran varietat de mètodes per representar i colorejar les molècules. També permet visualitzar la trajectòria de les dinàmiques moleculars de les proteïnes.

Aquest programa pot llegir arxius del PDB, els quals s'anomenen arxius pdb, i visualitzar l'estructura continguda. Aquest arxiu conté la posició de cada àtom, els aminoàcids que els àtoms formen, informació general de la proteïna, els autors responsables del descobriment de la proteïna i molts valors. Tot i així, l'únic necessari per a visualitzar la proteïna és saber la posició de cada àtom. Per a trobar el fitxer adequat al PDB, hem de saber el codi que el caracteritza o bé buscar la molècula que ens interessa i trobarem el codi amb la resta de la informació. Llavors ens hem de descarregar l'arxiu pdb i obrir-lo amb el VMD.

Els codis de les hemoglobines que estudiem són els següents:

Humà	Gos	Ànec de collverd	Colom	Tonyina
1A00	2QLS	3EOK	2R80	1V4U

## Imatges d'interès

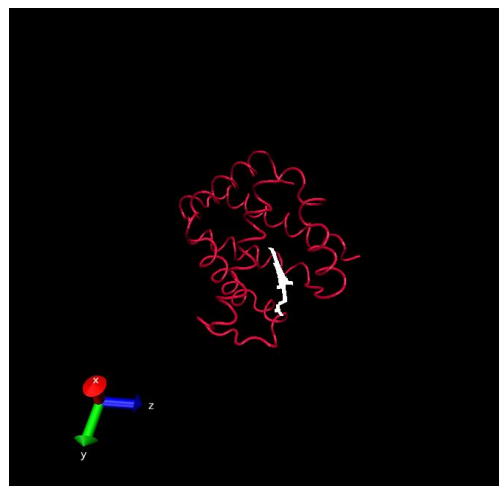
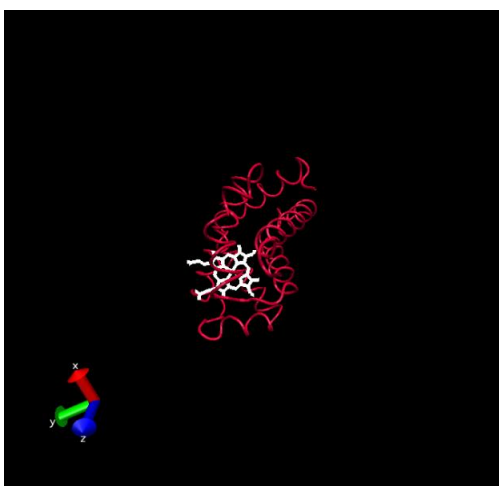
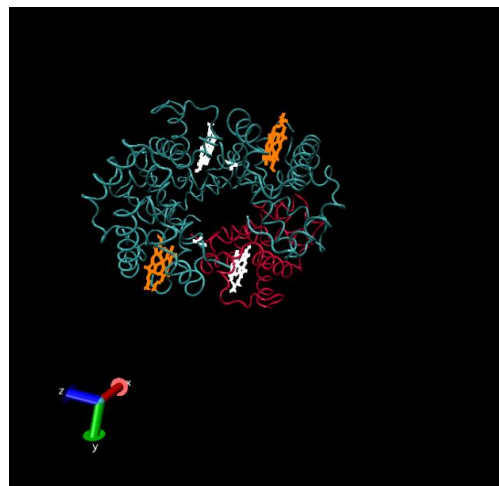
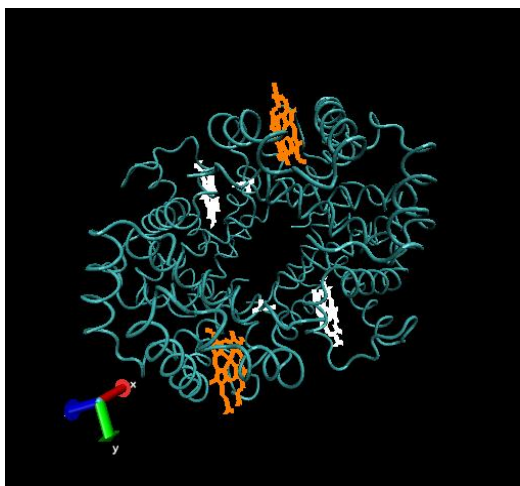
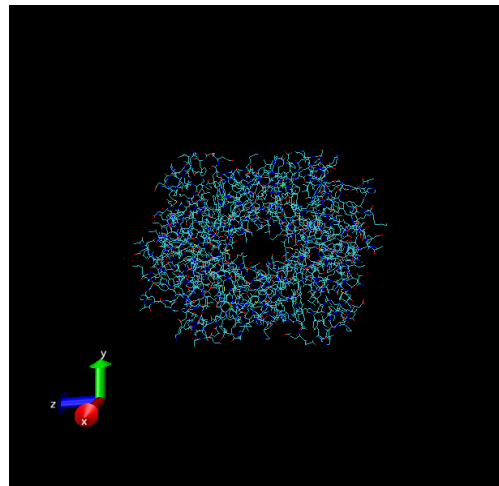
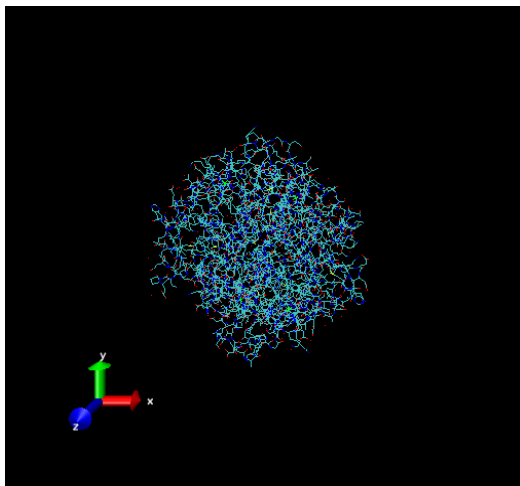
Per a començar a conèixer i comparar les diferents hemoglobines vam guardar sis imatges diferents:

- Dues perspectives de l'estructura quaternària de la molècula, o sigui, el conjunt de les diferents cadenes polipeptídiques que formen l'hemoglobina. Amb dues perspectives ens podem imaginar bé com és la molècula. La imatge és amb el *Drawing Method* (mètode de dibuix) a *Lines*, la qual cosa ens permet veure tots els enllaços en forma de línies.
- Una imatge amb els grups hemo ressaltats. Ara *Drawing Method* de les cadenes és *Tube*, el qual només resalta la cadena que uneix tots els aminoàcid, veient si l'estructura secundària és una hèlix  $\alpha$ , una làmina  $\beta$ ..., però no ens permet veure els residus. És amb el que treballarem normalment perquè ens permet tenir una visió més clara de l'estructura de la proteïna en general, ja que no ens molesten els residus.
- Una imatge igual que l'anterior però on trobem marcada de color vermell la cadena A, amb la qual treballarem.
- Dues perspectives de la cadena A.

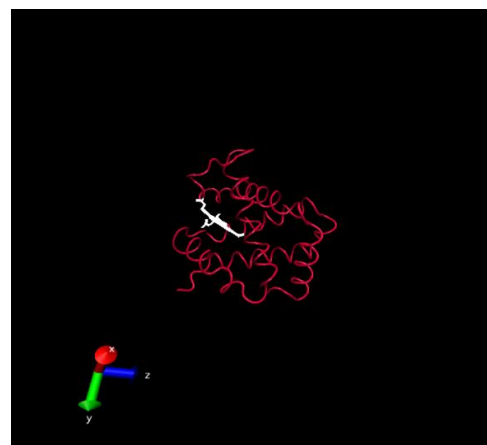
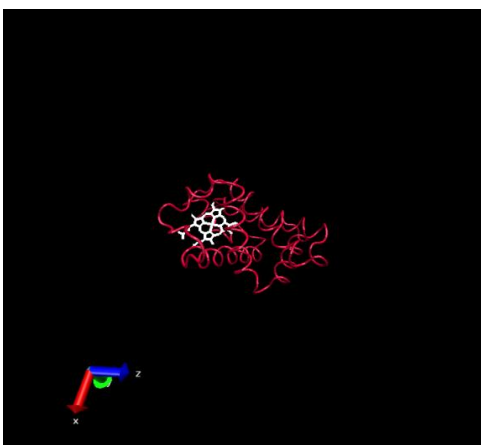
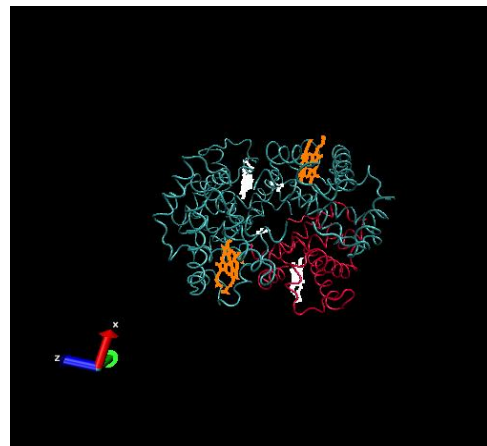
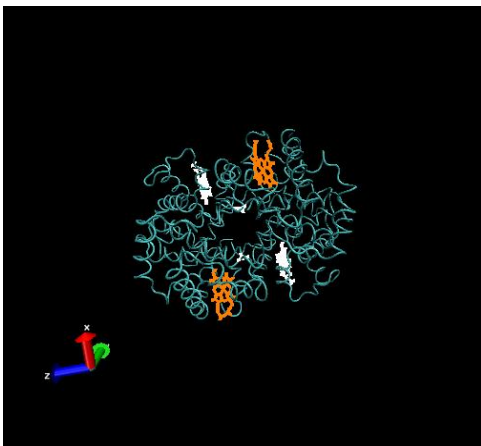
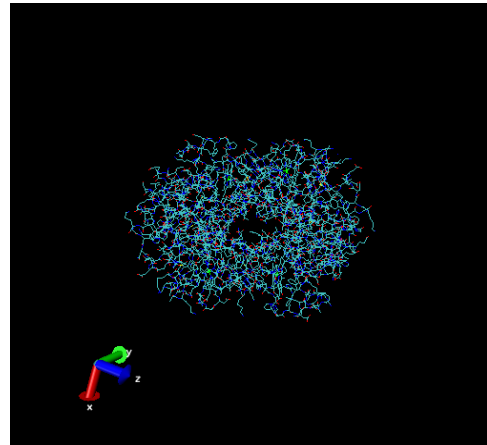
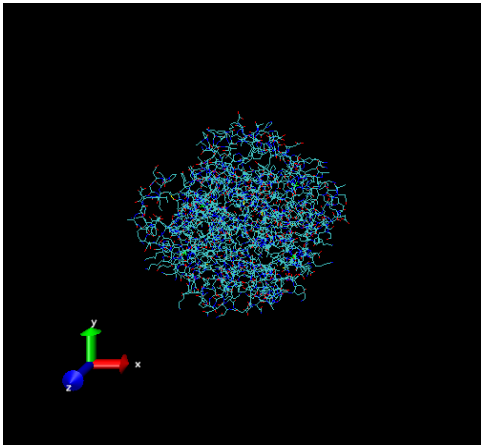


## Imatges obtingudes

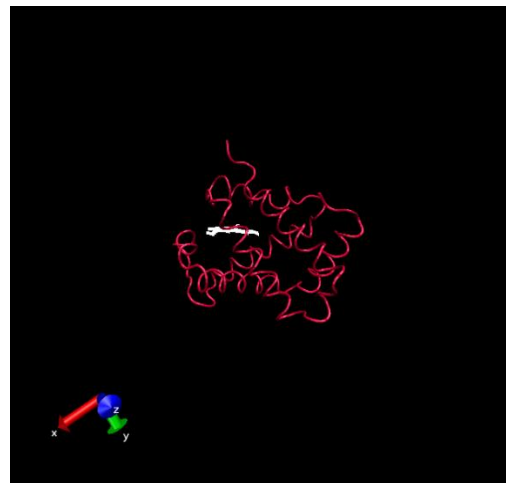
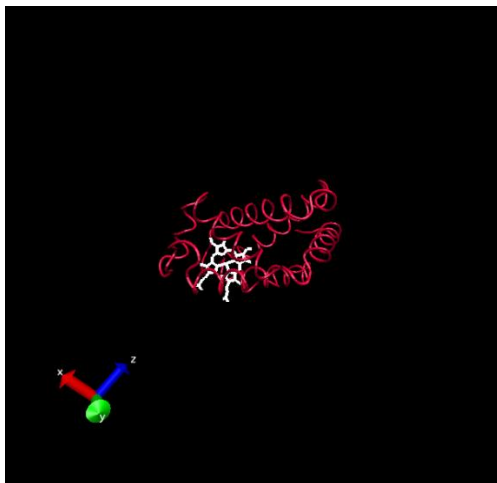
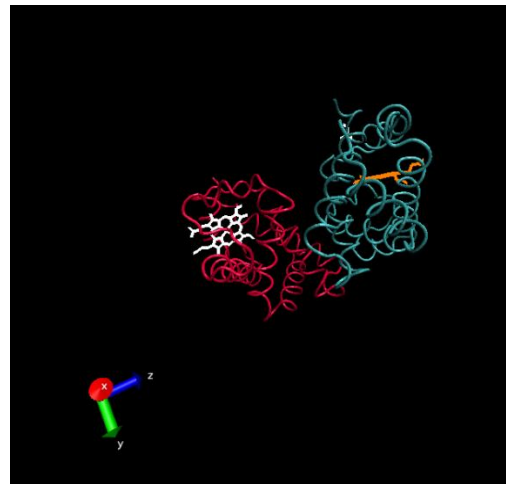
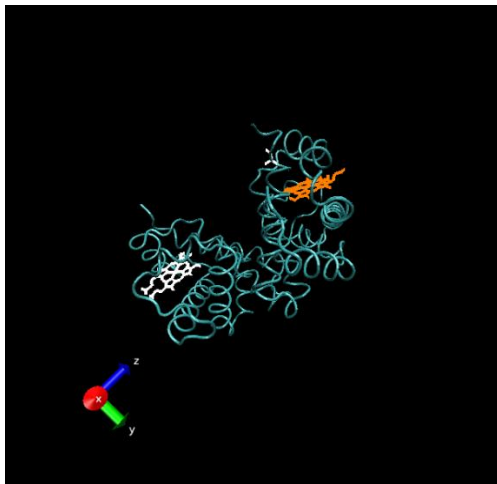
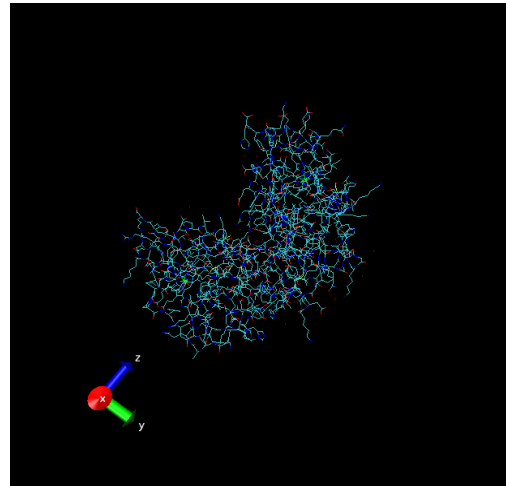
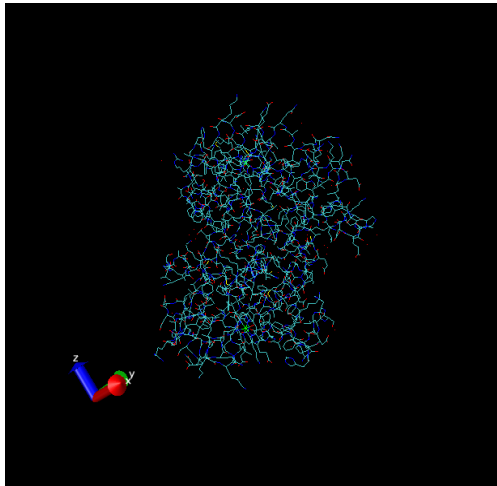
### Humà



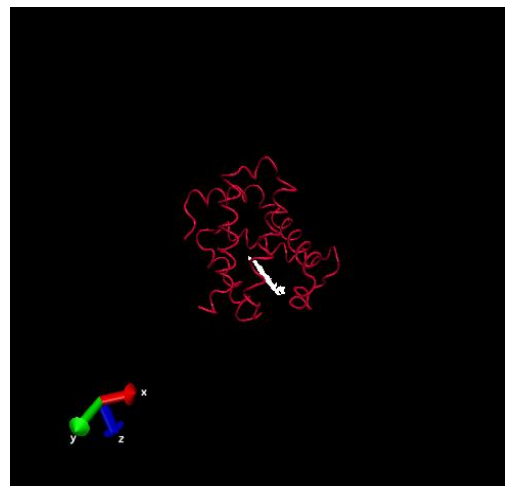
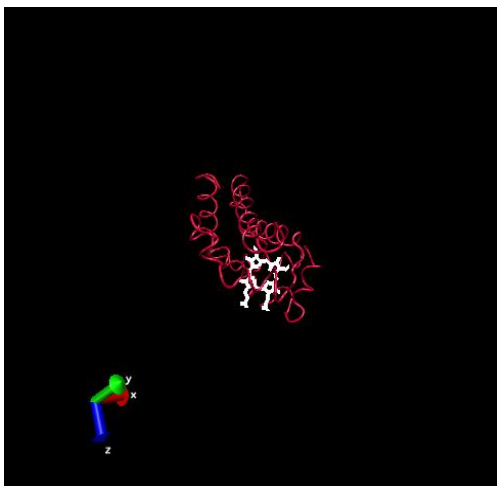
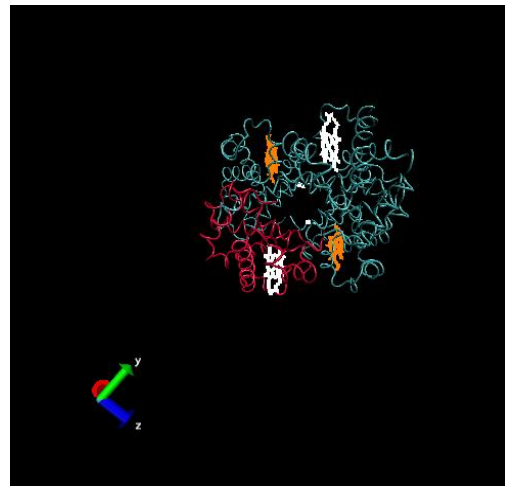
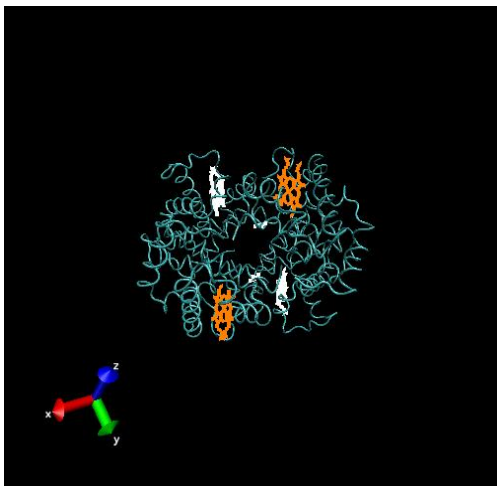
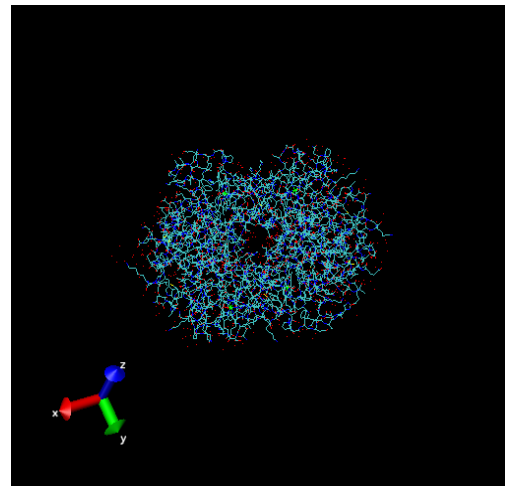
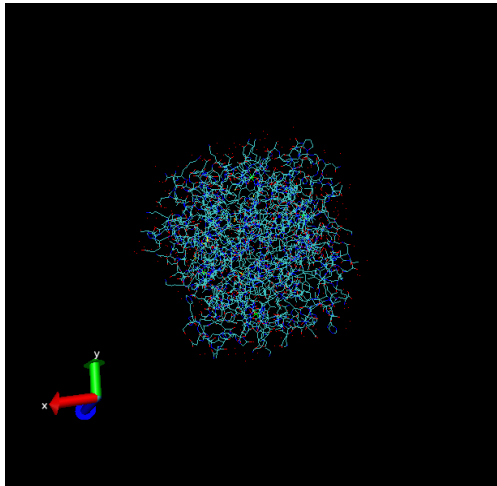
## Gos



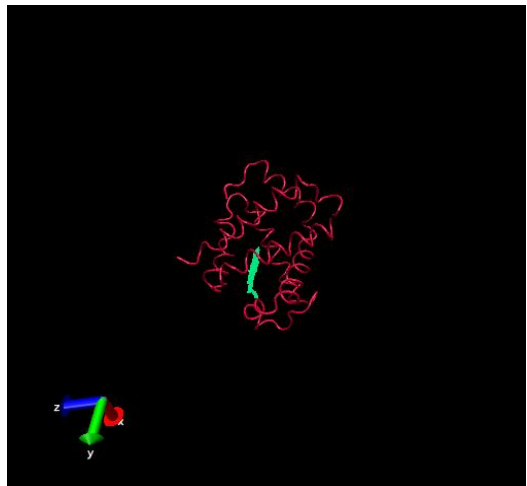
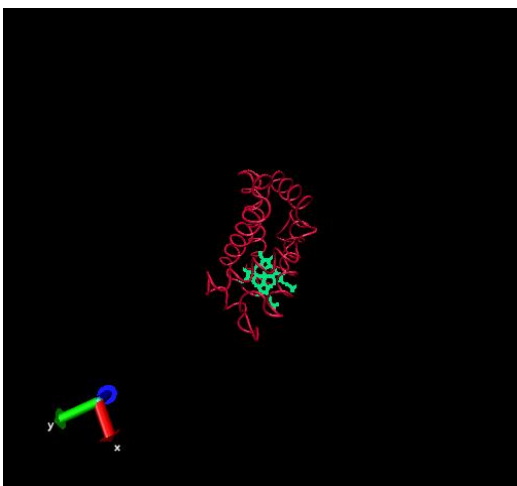
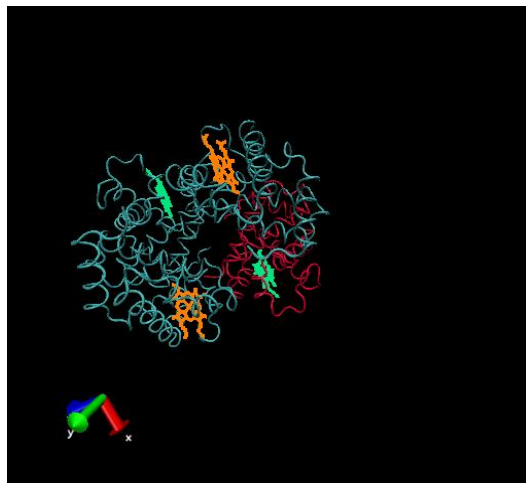
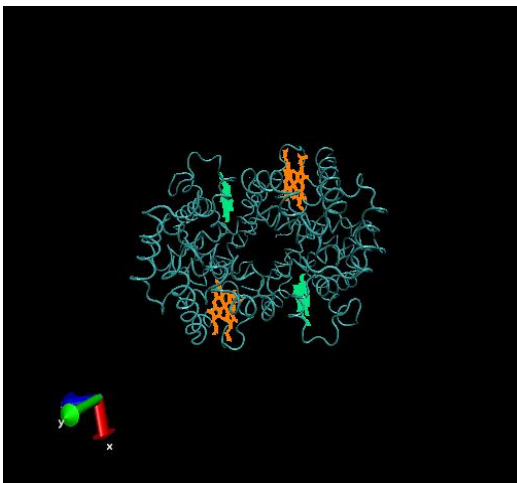
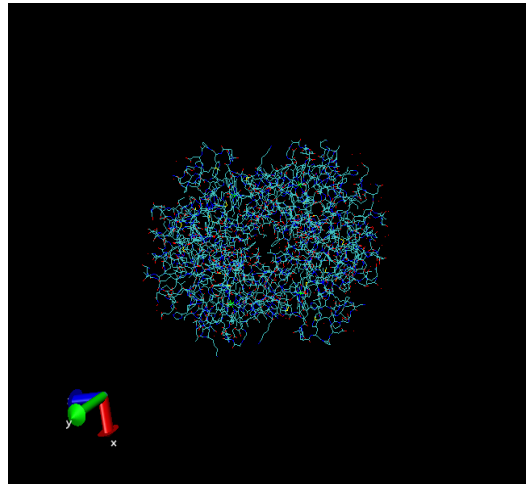
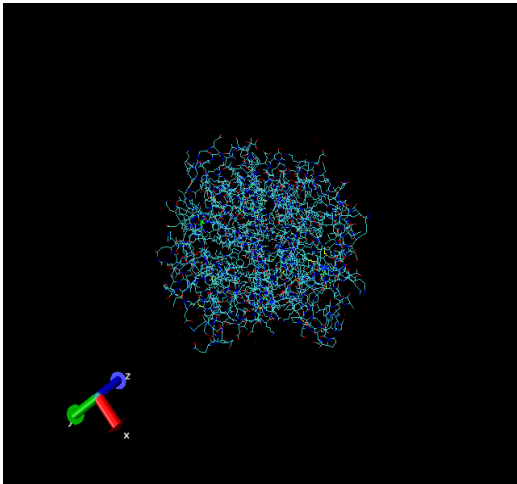
## Ànec de collverd



## Colom



## Tonyina



## Comentaris

El primer que ens va sobtar molt va ser veure que l'hemoglobina de l'humà, del gos, del colom i de la tonyina tenien quatre cadenes, mentre que l'ànec de collverd en tenia dues. Tot i que ja és possible que l'ànec de collverd només tingui dues cadenes, del que també es pot tractar és que el banc de dades del PDB només tingui la informació d'una cadena  $\alpha$  i una  $\beta$  de la proteïna. Aquesta pot ser una explicació, ja que en el cas del gos, vaig trobar al PDB una hemoglobina amb dues cadenes i una altra amb quatre cadenes. Vaig escollir la de quatre perquè és el cas més comú.

Per tal de ressaltar el grup hemo es necessita saber quin residu representa. Tot i no ser un aminoàcid, el programa VMD l'anomena residu. Vam veure que, efectivament, que trobàvem dos tipus de cadenes, on una era més curta i tenia el grup hemo al residu 142, mentre que l'altre el tenia al 147, demostrant que efectivament a l'hemoglobina hi trobem cadenes  $\alpha$  i  $\beta$ , on les cadenes  $\alpha$  estan formades per 141 aminoàcids i les  $\beta$  per 146. Per ressaltar aquest fet, els grups hemo de les cadenes  $\alpha$  són de color blanc i els de les cadenes  $\beta$ , taronja. Ara bé, mentre que les cadenes  $\beta$  són igual de llargues en totes les espècies, vam veure que el grup hemo en les cadenes  $\alpha$  de la tonyina era el residu 144. Representat de verd per aquest fet, ja ens indica que la cadena A és més llarga que la resta.

Centrant-nos en la cadena A de cada espècie, o sigui, les dues últimes imatges, podem veure que totes tenen una estructura molt similar independentment de l'espècie. A més, podem constatar que totes estan formades bàsicament per hèlix  $\alpha$ , les quals es troben en les mateixes posicions en totes les espècies. D'aquesta manera, per poder determinar amb molt més detall si les cadenes són diferents o no, hem de passar a comparar la seva estructura primària.

# Comparació de l'estructura primària de les diferents hemoglobines i establiment de relacions evolutives entre elles

## Material utilitzat

L'estructura primària (seqüència d'aminoàcids) de l'hemoglobina de cada espècie la vam extreure de la pàgina web del PCB, on l'anomenen *FASTA sequence*. D'allà vam obtenir la seqüència d'aminoàcids de totes les cadenes de l'hemoglobina, així que el que vam fer va ser seleccionar-ne la cadena A. Són les següents:

>**1A00**:A|PDBID|CHAIN|SEQUENCE (Humà)

VLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTN  
AVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSK  
YR

>**2QLS**:A|PDBID|CHAIN|SEQUENCE (Gos)

VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHFDLSPGSAQVKAHGKKVADALTAV  
AHLDDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFTPAVHASLDKFFAAVSTVLTSKYR

>**3EOK**:A|PDBID|CHAIN|SEQUENCE (Ànec de collverd)

VLSAADKTNVKGVSIGGHAAEEYGAETLERMFIAYPQTKTYFPHFDLSHGSAQIKAHGKKVAAALVEAVN  
HVDDIAGALSKLSDLHAQKLRVDPVNFKFLGHCFVVAIHHPAALTPEVHASLDKFMCAVGAVLTAKYR

>**2R80**:A|PDBID|CHAIN|SEQUENCE (Colom)

VLSANDKSNVKAFAKIGGQAGDLGGEALERLFITYPQTKTYFPHFDLSHGSAQIKGHGKKVAEALVEAAN  
HIDDIAGALSKLSDLHAQKLRVDPVNFKLLGHCFVVAHVHFPSSLTPEVHASLDKFLAVGTVLTAKYR


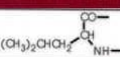
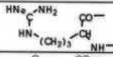
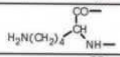
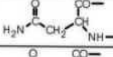
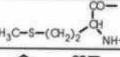
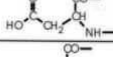
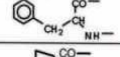


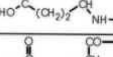

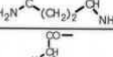


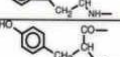
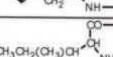
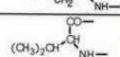
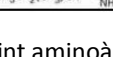
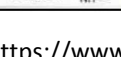
>**1V4U**:A|PDBID|CHAIN|SEQUENCE (Tonyina)

XTTSLDKDKSTVKALWGKISKADAIGADALGRMLAVYPQTKTYFSHWPDMSPGSGPVKAHGKKVMGG  
VALAVSKIDDLTTGLGDLSELHAFKMRVDPNSNFILSHCILVVVAKMFPKEFTPDHVS�DKFLASVALALA  
ERYR

Com es pot veure, en el *FASTA sequence* a cada aminoàcid li correspon una lletra i no tres com se sol veure. Aquesta nomenclatura és molt utilitzada quan es tracta de donar la seqüència de tota una proteïna, ja que així és molt més fàcil alinear i comparar diferents proteïnes. En la Imatge 3 es representen tots els aminoàcids i podem veure-hi les tres maneres d'anomenar-los.

Si mirem el principi de la seqüència de la tonyina, veiem que el primer és una X. Ara bé, si busquem aquesta lletra a la Imatge 3 veurem que aquesta lletra no correspon a cap aminoàcid. Així doncs, la lletra X s'utilitza quan no se sap quin aminoàcid és o quan no és rellevant quin sigui.

### Compositions for the Twenty Common Amino Acids

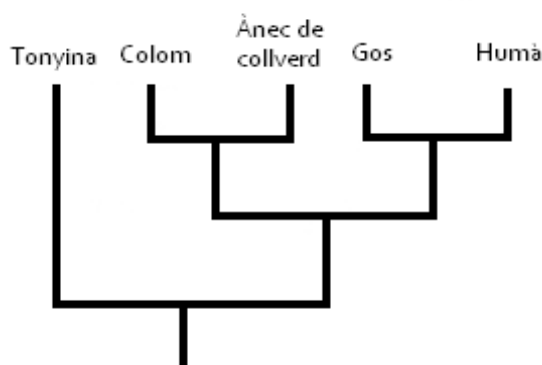
Name (Symbols)	Residue Composition	Residue Structure	Name (Symbols)	Residue Composition	Residue Structure
Alanine (Ala or A)	$C_3H_5NO$		Leucine (Leu or L)	$C_6H_{11}NO$	
Arginine (Arg or R)	$C_6H_{12}N_4O$		Lysine (Lys or K)	$C_6H_{12}N_2O$	
Asparagine (Asn or N)	$C_4H_8N_2O_2$		Methionine (Met or M)	$C_5H_9NOS$	
Aspartic Acid (Asp or D)	$C_4H_7NO_3$		Phenylalanine (Phe or F)	$C_9H_9NO$	
Cysteine (Cys or C)	$C_3H_5NOS$		Proline (Pro or P)	$C_5H_7NO$	
Glutamic Acid (Glu or E)	$C_5H_7NO_3$		Serine (Ser or S)	$C_3H_5NO_2$	
Glutamine (Gln or Q)	$C_5H_9N_2O_2$		Threonine (Thr or T)	$C_4H_7NO_2$	
Glycine (Gly or G)	$C_2H_3NO$		Tryptophan (Trp or W)	$C_{11}H_{10}N_2O$	
Histidine (His or H)	$C_6H_7N_3O$		Tyrosine (Tyr or Y)	$C_9H_9NO_2$	
Isoleucine (Ile or I)	$C_6H_{11}NO$		Valine (Val or V)	$C_5H_9NO$	

Imatge 3: Composició dels vint aminoàcids comuns (Font: <<https://www.msu.edu>>)

Per a comparar les seqüències més eficientment que a mà vam utilitzar un programa online anomenat ClustalW2, una eina del EBI (*European Bioinformatics Institut*, en anglès). Permet alinear múltiples seqüències d'ADN o proteïnes, calcula el millor aparellament per les seqüències seleccionades i les alinea per tal de mostrar les igualtats, semblances i diferències. Finalment dibuixa un arbre filogenètic per tal de veure les relacions evolutives entre totes les seqüències.

### Resultats esperats

Utilitzant la lògica, abans de recollir els resultats del programa ClustalW2 semblava clar que la seqüència d'aminoàcids més similar a l'humana havia de ser la del gos, ja que tots dos són mamífers. Seguidament, se sap que els mamífers tenen característiques morfològiques i fisiològiques més semblants als ocells que als peixos: a diferència dels peixos, ocells i mamífers posseeixen quatre extremitats, respiració pulmonar,... Així que el gos i l'humà havien d'estar més a prop del colom i de l'ànec de collverd que de la tonyina. I, com que no hi havia cap raó que ens pogués fer creure que un dels dos ocells estava més a prop evolutivament als mamífers que l'altre, vam imaginar que tots dos haurien d'estar relacionats evolutivament al gos i a l'humà de la mateixa manera. Aquest raonament es pot veure representat en l'arbre filogenètic de la Imatge 4.



Imatge 4: Arbre filogenètic esperat de les cinc espècies estudiades (Font: Dibuix realitzat amb Paint)



Segons els resultats esperats, les nostres cinc espècies tenien un avantpassat en comú que, després d'un procés d'especiació, va donar lloc a dues espècies, una de les quals acabaria sent la tonyina de manera directa i l'altra donaria lloc a les altres espècies després de dos processos d'especiació més: en el primer obtindríem dues espècies que representarien els dos ocells i els dos mamífers respectivament, i en el segon ja trobaríem el colom i l'ànec de collverd, i el gos i l'humà. Naturalment, l'avantpassat de les cinc espècies va donar lloc a moltes més espècies i, per tant, una de les espècies resultant del primer procés d'especiació segur que no va donar lloc només a la tonyina. Tot i així, un arbre filogenètic és tan sols un model i, per tant, simplifica la realitat, quedant-se amb només les cinc espècies d'interès, en el nostre cas.

Un fet molt important és no caure en l'error de l'"escala de la vida". Des del temps dels grecs antics que es pensava en la natura com a una escala de la vida. A cada graó hi havia grups d'animals com "insectes", "serps", "peixos", i els graons simbolitzaven que una espècie provenia de l'espècie del graó de sota. A baix de tot s'hi trobaven els quatre elements (foc, aire, aigua i terra) i, a mesura que s'anava pujant, anava augmentant el grau de perfecció, amb l'ésser humà a dalt de tot de l'escala, indicant que era el més perfecte de tot el que hi ha a la natura.

Darwin va rebutjar aquest model i va defensar el model d'arbre, el dels arbres filogenètics. De fet, l'única figura que apareix al seu gran llibre *l'Origen de les espècies* és un arbre filogenètic. Ell va tenir la brillant idea de proposar que tots els éssers vius provenim d'un avantpassat comú. Per tant, hi ha una unitat fonamental de vida. A més a més, la biologia evolutiva també ens ha permès saber que les espècies no són estàtiques, sinó que el seu patrimoni genètic canvia.

Els anys següents a les obres de Darwin, els biòlegs van començar a donar per bona la forma d'arbre. Tot i així, encara ara se senten restes del pensament de l'escala de la vida. Quans cops hem sentit allò de "l'home prové del mico"? En aquesta afirmació s'entendria que els micos són un avantpassat dels humans, cosa que no és veritat. El que sí és cert és que l'avantpassat comú de micos i humans era més semblant als micos que als humans, però no és correcte dir que aquest avantpassat fos un mico.

Per tant, tornant al nostre arbre filogenètic, no podem dir que els ocells ni els mamífers provenen dels peixos. Els nostres avantpassats, els dels humans, no eren ni les tonyines, ni els coloms ni els ànecs de collverd. És el mateix cas que el dels humans i els micos.

## Resultats obtinguts

### Alineació

```
1A00_A|PDBID|CHAIN|SEQUENCE      --VLSPADKTNVKAAWGKVGAGHAGEYGAELERMFLSFPTTKTYFPHF-D 47
2QLS_A|PDBID|CHAIN|SEQUENCE      --VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPTTKTYFPHF-D 47
3EOK_A|PDBID|CHAIN|SEQUENCE      --VLSAADKTNVKGVSFKIGGHAEYGAETLERMFIAYPQTKTYFPHF-D 47
2R80_A|PDBID|CHAIN|SEQUENCE      --VLSANDKSNVKAVFAKIGGQAGDLGGEALERLFTYPQTKTYFPHF-D 47
1V4U_A|PDBID|CHAIN|SEQUENCE      XTTLSDKDKSTVKALWGKISKSDAIGADALGRMLAVYPQTKTYFSHWPD 50
                                   .** **:.:.: * : *:. * *:.: * * : : * *****.: *

1A00_A|PDBID|CHAIN|SEQUENCE      LSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVN 97
3N48_A|PDBID|CHAIN|SEQUENCE      LSPGSAQVKAHGKKVADALTAVAHLDLPGALSALSDLHAYKLRVDPVN 97
3EOK_A|PDBID|CHAIN|SEQUENCE      LSHGSAQIKAHGKKVAAALVEAVNHVDDIAGALSKLSDLHAQKLRVDPVN 97
2R80_A|PDBID|CHAIN|SEQUENCE      LSHGSAQIKGHGKKVAEALVEAANHIDDIAGALSKLSDLHAQKLRVDPVN 97
1V4U_A|PDBID|CHAIN|SEQUENCE      MSPGSGPVKAHGKKVMGVALAVSKIDDLTTGLGDLSELHAFKMRVDPSPN 100
                                   : * ** . : *.***** .:. * . :.:.: . . . * :.*** * :.*** *

1A00_A|PDBID|CHAIN|SEQUENCE      FKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR 141
3N48_A|PDBID|CHAIN|SEQUENCE      FKLLSHCLLVTLACHHPTEFTPAVHASLDKFFAAVSTVLTSKYR 141
3EOK_A|PDBID|CHAIN|SEQUENCE      FKFLGHCFLLVVVAIHHPAALTPEVHASLDKFMCAVGAVLTAKYR 141
2R80_A|PDBID|CHAIN|SEQUENCE      FKLLGHCFLLVVAVHFPSSLTPEVHASLDKFLAVGTVLTAKYR 141
1V4U_A|PDBID|CHAIN|SEQUENCE      FKILSHCILVVAKMFPKEFTPAHVSLDKFLASVALALAERYR 144
                                   **:*.**:*.**: * : ** .*.***** .:*. .*: : **
```

Amb aquesta aplicació aconseguim comparar les cinc cadenes aminoàcid per aminoàcid. Sota de les cinc seqüències d'aminoàcids hi trobem una fila amb diversos símbols. Cada símbol, o l'absència d'aquests, correspon a la columna d'aminoàcids que té a sobre. Un "\*" significa que els aminoàcids són idèntics en totes les seqüències; una conservació total. Els ":" i "." indiquen diferents graus de conservació: els dos punts indiquen que hi ha hagut substitucions, però que aquestes són conservatives, mentre que el punt significa que la substitució és semi conservativa. El fet que se n'utilitzi un o l'altre depèn de les propietats físiques i químiques dels aminoàcids. L'absència de símbol indica que les substitucions no tenen cap grau de conservació.

Com havíem dit abans, veiem que la cadena de la tonyina és més llarga. Ara bé, sorprenentment és tres aminoàcids més llarga, com podem veure en els números del final de tot de la seqüència, i no dos com ens pensàvem després dels resultats del VMD. L'única explicació lògica és que el VMD no tingui en compte l'aminoàcid desconegut, representat amb una X, ja que no el pot representar. Per tant, pel VMD, la seqüència d'aminoàcids de la tonyina comença en la primera treonina (T). La posició dels aminoàcids extra de la tonyina és la següent: els dos primers es troben al principi de tot i el tercer a la posició 49 de la cadena de la tonyina, el qual es trobaria entre els aminoàcids 46 i 47 de les altres cadenes. Això fa que trobem gaps (forats) en aquests tres punts de les seqüències dels ocells i dels mamífers.

Pel que fa a la conservació, podem destacar zones altament conservades, on fragments de la seqüència són idèntics en totes les espècies. Tot i així, la majoria de l'alineament ens mostra que predominen les zones amb semi conservació o gens de conservació. Amb només l'alineament, però, no sabem si és una seqüència que difereix molt de les altres la que provoca aquest resultat, o bé si totes difereixen igual de totes. Per tal de saber-ho amb precisió, hem de mirar la taula de puntuacions de continuació.

## Taula de puntuacions

SeqA	Name	Len (aa)	SeqB	Name	Len (aa)	Score%
1	1A00_A PDBID CHAIN SEQUENCE	141	2	2QLS_A PDBID CHAIN SEQUENCE	141	<b>83</b>
3	3EOK_A PDBID CHAIN SEQUENCE	141	4	2R80_A PDBID CHAIN SEQUENCE	141	<b>82</b>
1	1A00_A PDBID CHAIN SEQUENCE	141	3	3EOK_A PDBID CHAIN SEQUENCE	141	<b>71</b>
2	2QLS_A PDBID CHAIN SEQUENCE	141	3	3EOK_A PDBID CHAIN SEQUENCE	141	<b>70</b>
2	2QLS_A PDBID CHAIN SEQUENCE	141	4	2R80_A PDBID CHAIN SEQUENCE	141	<b>69</b>
1	1A00_A PDBID CHAIN SEQUENCE	141	4	2R80_A PDBID CHAIN SEQUENCE	141	<b>68</b>
1	1A00_A PDBID CHAIN SEQUENCE	141	5	1V4U_A PDBID CHAIN SEQUENCE	144	<b>55</b>
2	2QLS_A PDBID CHAIN SEQUENCE	141	5	1V4U_A PDBID CHAIN SEQUENCE	144	<b>53</b>
4	2R80_A PDBID CHAIN SEQUENCE	141	5	1V4U_A PDBID CHAIN SEQUENCE	144	<b>52</b>
3	3EOK_A PDBID CHAIN SEQUENCE	141	5	1V4U_A PDBID CHAIN SEQUENCE	144	<b>51</b>

En aquesta taula es relacionen totes les seqüències per parelles. Així, en cada fila tenim dues seqüències amb el seu nombre d'aminoàcids i, al final a *score*, els aminoàcids que tenen en comú les dues seqüències. Estan ordenades segons l'*score* de cada combinació, de gran a petit.

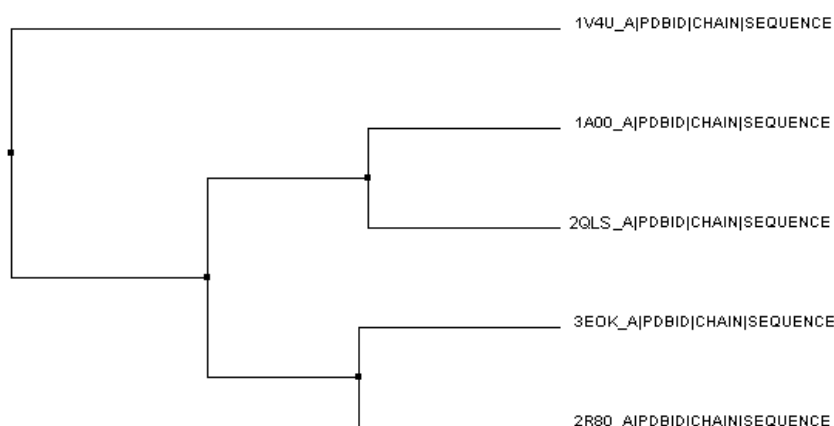
Per a calcular l'*score* primer s'agafa el nombre d'aminoàcids coincidents en les dues seqüències i es divideix pel nombre total de residus comparats. Llavors només queda passar-lo a tant per cent. D'aquesta manera l'*score* ens indica el tant per cert d'identitat (parts idèntiques) entre les dues seqüències.

El que ressalta més és que en les quatre combinacions amb l'*score* més baix, una de les dues seqüències és sempre la de la tonyina. Aquest fet ens indica que l'hemoglobina de la tonyina és la que difereix més de la resta d'hemoglobines. Per tant, la causant dels resultats poc conservats de l'alineament era la tonyina. Si mirem en més detall els valors de l'*score* en cada cas, veurem un factor que podria semblar sorprenent: la tonyina té un tant per cent d'identitat més alt amb els mamífers que no pas amb els ocells. Ens podria semblar que els resultats han de ser erronis dos motius:

1. Si utilitzem la intuïció i, per tant, el pensament de l'"escala de la vida", l'hemoglobina de la tonyina s'hauria d'assemblar més a la dels ocells que no pas a la dels mamífers. Això s'explica veient que els peixos i els ocells estan en graons més pròxims que no pas els peixos i els mamífers, ja que es pensa que un avantpassat dels ocells van ser els peixos i un avantpassat dels mamífers van ser els ocells i, encara més enrere, els peixos. Tot i així, ja hem dit que aquest pensament no és vàlid i, per tant, aquest possible argument per donar per incorrectes els resultats ha de quedar clar que no és bo.
2. Pel que esperem i que hem representat a l'arbre, els ocells i els mamífers estan igual d'allunyats evolutivament de la tonyina perquè tenen un avantpassat comú. Tot i així, això no treu que potser la branca dels ocells hagi sofert més mutacions en el gen de l'hemoglobina que la dels mamífers i que, per això, difereixi més de l'hemoglobina de la tonyina.

També podem destacar que el tant per cent d'identitat entre mamífers i ocells en les quatre combinacions possibles és molt semblant i, per tant, això demostra que els dos mamífers estan igualment relacionats evolutivament amb els dos ocells, i no amb un més que amb l'altre, tal i com havíem pensat. Per últim, els mamífers entre ells i els ocells entre ells tenen un alt tant per cent d'identitat, superior al 80%, cosa que ens indica que estan molt a prop evolutivament. Pel que sembla, els nostres resultats esperats són correctes.

## Arbre filogenètic



Aquí podem veure l'arbre filogenètic dibuixat pel programa ClustalW2. Per ser capaços de comparar aquest arbre amb el que havíem suposat a "Resultats esperats", hem de recordar la teoria dels arbres filogenètics. Podem veure que aquest arbre té la mateixa forma que el nostre, però el grup dels mamífers i els ocells s'han intercanviat el lloc, igual que l'humà i el gos ho han fet dins el grup dels mamífers, i l'ànec de collverd i el colom en el grup dels ocells. Com hem dit anteriorment, això no és un problema perquè les branques dels arbres poden rotar, sempre i quan aquestes continuïn naixent del mateix node. Per tant, el nostre arbre i aquest ens aporten exactament la mateixa informació i, com que els resultats esperats han acabat sent els obtinguts, podem concloure que la nostra hipòtesi era correcta.

El que s'ha de comentar és que no es pot assegurar que, agafant un altre gen, no ens sortís un arbre diferent. O sigui, que posar la mà al foc que aquesta relació que mantenen les proteïnes de les diferents espècies és la mateixa entre les espècies en general és delicat. Però bé, encara que no es pugui assegurar el 100%, els resultats són prou de fiar.

### ***Explicació de l'alineació***

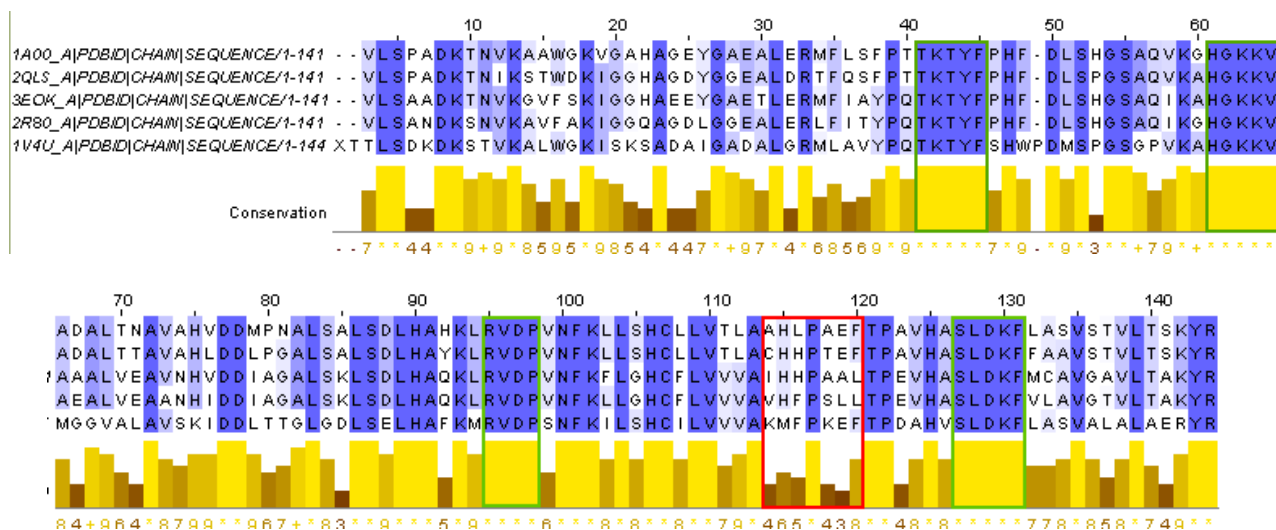
Pel que fa a l'alineació, ara que sabem que aquest és l'arbre filogenètic correcte, podem explicar quan es devien produir les mutacions necessàries que expliquen que la tonyina tingui 144 aminoàcids i la resta només 141. Com que els punts on trobem els *gaps* són els mateixos en totes les hemoglobines amb 141 aminoàcids, podem formular dues teories. Podem pensar que després de la primera especiació, l'espècie que havia de donar lloc a ocells i mamífers va patir un mínim de dues deleccions abans de sofrir la següent especiació: una devia provocar la deleció dels dos primers aminoàcids i, per tant, es devia produir una deleció dels sis primers nucleòtids del gen que codifica l'hemoglobina; l'altra devia provocar la deleció del 49è aminoàcid de la proteïna inicial, causada per la deleció dels nucleòtids que el codificaven. L'altra opció és que la tonyina patís un mínim de dues insercions, o sigui, exactament que el cas contrari però substituint les deleccions per insercions.

### ***Explicació de la taula de puntuacions***

Aquí queda representat, també, els resultats que podíem veure en la taula de puntuacions. Primer de tot, veiem que efectivament la tonyina era la que diferenciava de la resta d'hemoglobines. Segonament, és lògic tenint l'arbre filogenètic al davant que l'hemoglobina del gos i de l'humà, i l'hemoglobina del colom i l'ànec de collverd tinguin un tant per cent d'identitat tan alt en els seus respectius casos. Per últim, es veu per què el tant per cent d'identitat entre les seqüències d'aminoàcids d'ocells i mamífers era molt més alt que no pas entre aquests grups i la tonyina, ja que ocells i mamífers tenen un avantpassat comú posterior a l'avantpassat comú que tenen amb la tonyina.

# Estudi de la zona menys conservada i les zones perfectament conservades

## Zones estudiades



El programa ClustalW2 també permet visualitzar els resultats amb una aplicació de Java, la qual s'anomena JalView. Es poden visualitzar les seqüències estudiades i, ordenar-les de la manera que es prefereixi i pintar-les ressaltant el que ens interressi. En el cas de dalt, els aminoàcids estan pintats amb diferents tons de blau segons el grau de conservació de l'aminoàcid per cada columna. Així, les columnes perfectament conservades són d'un blau intens i les gens conservades són blanques.

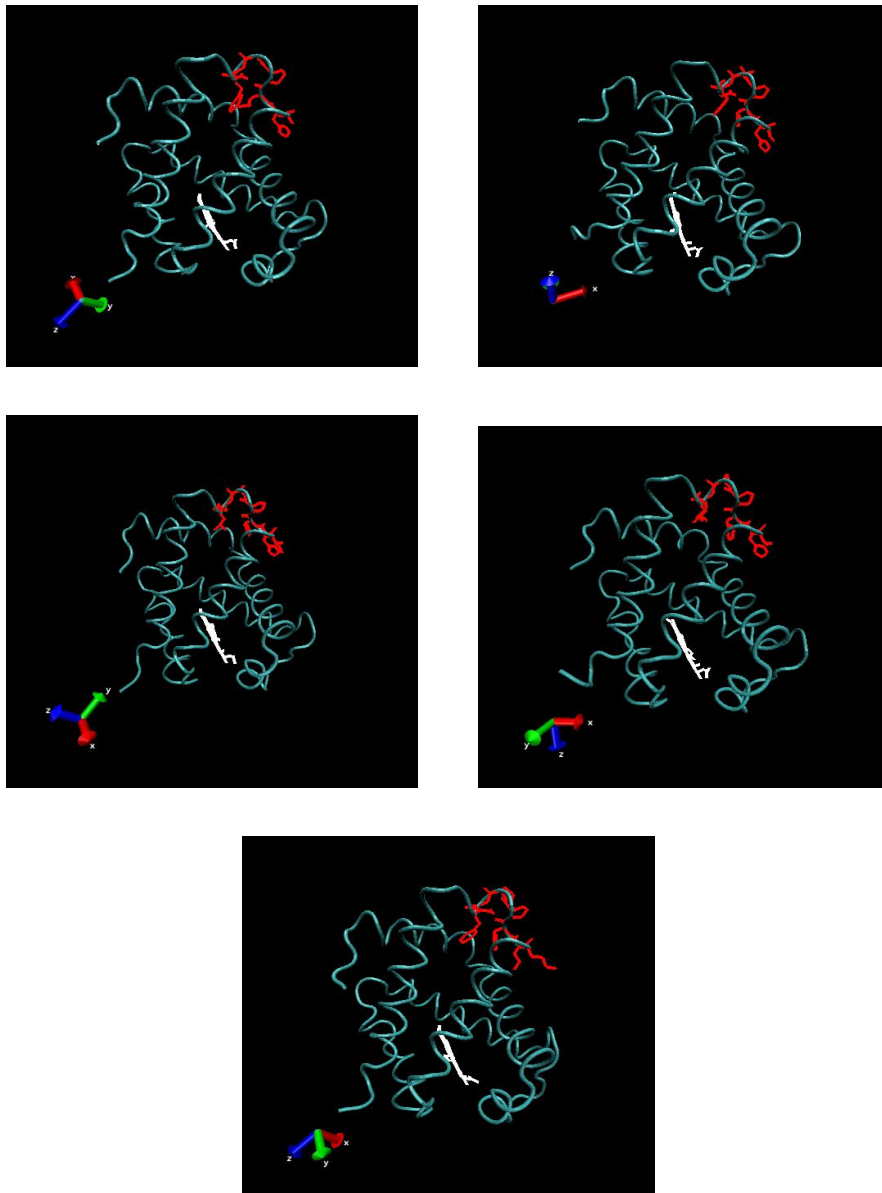
A més, a sota de les seqüències veiem que també s'hi representa la conservació de cada columna d'aminoàcids. En aquest cas, es fan servir diferents tons que van de groc (perfectament conservat) fins a marró (molt poc conservades). En aquest apartat estudiarem la zona menys conservada (marcada amb un rectangle vermell) i les zones perfectament conservades (marcades amb rectangles verds).

En el subapartat de la zona menys conservada, l'objectiu serà veure en quins punts de l'arbre filogenètic es van produir les mutacions que van comportar la gran varietat d'aminoàcids en un mateix punt de les diferents seqüències, tenint sempre en compte el principi de la mínima evolució i la màxima parsimònia. L'arbre filogenètic utilitzat és l'obtingut amb el ClustalW2. En el subapartat de les zones perfectament conservades, pot ser interessant veure la situació d'aquestes zones en l'estructura terciària de les hemoglobines, per tal de trobar una explicació a l'alta conservació.

## Situació de la zona menys conservada en l'estructura terciària i situació de les mutacions en l'arbre filogenètic obtingut

### Situació de la zona menys conservada en l'estructura terciària

La zona estudiada comprèn els aminoàcids 111-117 d'ocells i mamífers i els aminoàcids 114-120 de la tonyina. No són zones diferents, sinó que s'ha de tenir en compte que la tonyina, en aquest punt, té tres aminoàcids més que la resta. Per situar aquestes zones a l'estructura terciària hem de tenir en compte que el VMD no detecta el primer aminoàcid de la tonyina. Per tant, haurem de situar 111-117 per a ocells i mamífers i 113-119 per a la tonyina. Aquestes zones les trobem ressaltades de vermell. Les imatges corresponen, per aquest ordre, a l'hemoglobina d'humà, de gos, d'ànec de collverd, de colom i de tonyina.



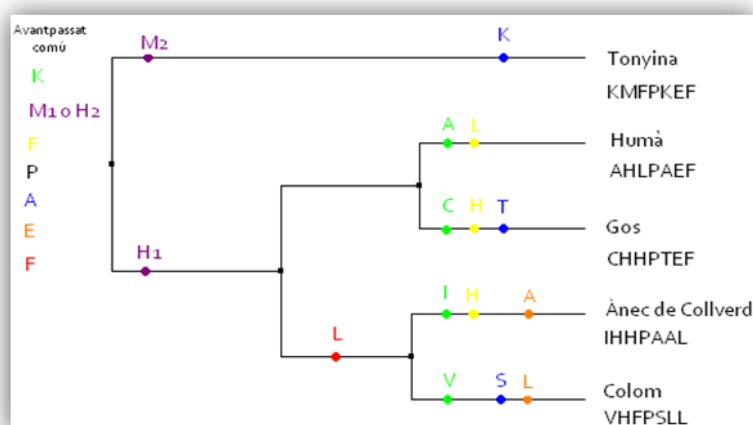
En aquestes imatges veiem clarament que la zona menys conservada se situa, en totes les hemoglobines, a la zona més allunyada del grup hemo, cosa que explica que s'hagi conservat tan poc.

## Situació de les mutacions produïdes a la zona menys conservada en l'arbre filogenètic obtingut

Estudiarem set columnes, i explicarem el que devia passar en cada una per separat. Per entendre els raonaments és molt important tenir present a quina espècie pertany cada fila, i mirar quin aminoàcid correspon a cada espècie abans de llegir el raonament. Cada número de la llista numerada següent representa la columna de la qual parlem. O sigui, el número 1 indica que parlem de la primera columna. A més, al principi de cada punt hi ha indicat un color, el qual ens servirà per identificar les mutacions en l'arbre filogenètic que segueix la llista numerada.



1. Verd. A la primera columna tots els aminoàcids són diferents. Per tant, no és possible saber quin devia ser l'aminoàcid que tenia l'avantpassat comú de les cinc espècies. Hem agafat la K perquè la tonyina és la que s'acosta més a l'avantpassat, però podria ser qualsevol.
2. Lila. Tots tenen una Histidina menys la tonyina, que té una metionina. Tenim dues possibilitats igual de provables perquè només necessitem una mutació: o bé l'avantpassat comú tenia una M i llavors la branca mare dels ocells i mamífers va patir una mutació, o bé l'avantpassat comú tenia una H i la tonyina va patir una mutació.
3. Groc. Considerarem que l'avantpassat comú tenia una F perquè, tot i que la tonyina i el colom estan bastant lluny evolutivament, coincideixen amb aquest aminoàcid. Tot i que el gos i l'ànec de collverd comparteixen la H, no ens queda més remei que pensar que aquest fet va ser causat per dues mutacions diferents que van acabar donant lloc al mateix resultat.
4. La quarta columna té una conservació perfecta: totes les espècies tenen una prolina en aquest punt. Per tant, tenim clar que l'avantpassat comú tenia una P i que no s'ha fixat cap mutació en aquest punt al llarg del temps.
5. Blau. En aquest cas també tenim molta variabilitat, però no suposarem que l'avantpassat comú tenia una K com la tonyina perquè llavors necessitaríem quatre mutacions per explicar els altres aminoàcids, mentre que si suposem que tenia una A com l'humà i l'ànec de collverd, només en necessitem tres.
6. Taronja. En aquest cas és bastant clar que l'avantpassat comú tenia una E, i que només es van produir mutacions dins el grup dels ocells.
7. Vermell. L'avantpassat devia tenir molt probablement una F, i una mutació es va produir a la branca mare dels ocells, ja que tots dos tenen una L.



S'ha de dir que només importa en quina branca està situada la mutació, però no a quina alçada de la branca. Per exemple, en la branca de l'humà, la mutació pintada de verd no té per què haver ocorregut abans de la groga. Simplement la verda va primera perquè hi estan representades per ordre de columna.

Un cop realitzada la situació de les possibles mutacions, podia ser interessant veure si aquests canvis d'aminoàcids eren fàcils de produir-se o no, dit d'una altra manera, comprovar si amb una simple mutació puntual es podia passar d'un aminoàcid a un altre de la mateixa columna. No es van considerar possibles deleccions o insercions perquè aquestes haguessin fet variar tota la continuació de la proteïna, i tampoc casos de deleccions i insercions que es complementen perquè l'objectiu era veure un canvi fàcil de produir-se, com una mutació puntual, i no casos tan rebuscats. La manera de comprovar-ho era consultant la taula del codi genètic.

### Comprovació de la viabilitat de les mutacions estudiades consultant el codi genètic

A part de veure si amb una mutació puntual podem obtenir un altre aminoàcid de la columna, el que també volem aconseguir fent aquest estudi és veure s'han de retocar les suposicions de l'apartat anterior sobre en quin punt evolutiu es devien produir els mutacions. El motiu és que potser veurem més lògic que l'avantpassat comú tingués un altre aminoàcid, cosa que farà necessaris canvis posteriors.

La taula del codi genètic utilitzada (Imatge de la dreta) és la mateixa que apareix a l'apartat teòric de mutacions, només que està modificada per tal que també hi aparegui l'altra manera d'anomenar els aminoàcids, amb una sola lletra, ja que és la nomenclatura que nosaltres sempre utilitzem.

		Second nucleotide				
		U	C	A	G	
First nucleotide	U	UUU Phe F UUC UUA Leu L UUG	UCU UCC Ser S UCA UCG	UAU Tyr Y UAC UAA STOP UAG STOP	UGU Cys C UGC UGA STOP UGG Trp W	U C A G
	C	CUU CUC Leu L CUA CUG	CCU CCC Pro P CCA CCG	CAU His H CAC CAA Gln Q CAG	CGU CGC CGA CGG	U C A G
	A	AUU Ile I AUC AUA AUG Met M	ACU ACC ACA Thr T ACG	AAU Asn N AAC AAA Lys K AAG	AGU Ser S AGC AGA AGG Arg R	U C A G
	G	GUU GUC Val V GUA GUG	GCU GCC Ala A GCA GCG	GAU Asp D GAC GAA Glu E GAG	GGU GGC GGA GGG	U C A G

També analitzarem cada columna per separat. Igual que abans, el número de la llista ens indica de quina columna parlem i el color es manté:

1. Verd. Hi trobem els aminoàcids A, C, I, V i K. Després de mirar a la taula, es veu que amb una sola mutació puntual només es pot fer el canvi I-V (AUU-GUU i AUC-GUC), per descomptat que sempre pels dos sentits, i A-V (GUU-GCU, GUC-GCC, GUA-GCA i GUG-GCG) Aquest canvi és més probable que el primer perquè hi ha més possibles mutacions puntuals. Pels altres es necessiten més d'una mutació puntual.

Ara podríem plantejar-nos el que teníem pensat inicialment: abans hem col·locat una K a l'avantpassat comú, però hem dit que podia ser qualsevol altre. Després de saber els possibles canvis que s'haurien pogut produir, és millor dir que l'avantpassat tenia una V, ja que a partir d'aquesta podem obtenir la A i la I de l'humà i l'ànec de collverd, respectivament, ja que si la V ve codificada pels nucleòtids GUU, amb un canvi obtenim AUU (I) o bé GCU (A). A més, el pas de K a C (l'aminoàcid que té el gos) és menys probable que el de V a C, ja que implica canviar completament els tres nucleòtids, mentre que la segona possibilitat només necessita dos canvis.

2. Lila. Només hi tenim la M i la H. Els codons que codifiquen per la metionina són AUA i AUG, i per la Histidina són CAU i CAC. Com es pot veure, amb una sola mutació



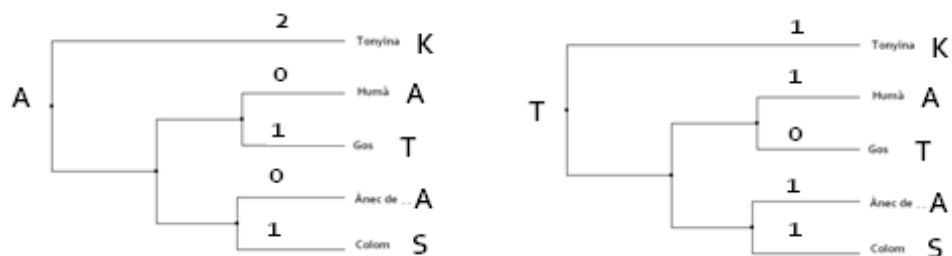
puntual no podem passar d'una a l'altra. De fet, necessitaríem un canvi en els tres nucleòtids del codó. Del que teníem pensat, no cal canviar res perquè continuen sent igual de probables les dues possibles mutacions que havíem dit.

3. Groc. Hi trobem L, H i F. En aquest cas hi ha una gran quantitat de possibles canvis, ja que és possible el canvi L-H i L-F amb una sola mutació puntual. Pel canvi L-H tenim CUU-CAU i CUC-CAC, i pel canvi L-F tenim UUA o UUG per UUU o UUC i a l'inrevés, o bé CUU-UUU i CUC-UUC. Com en la primera columna, el canvi L-F també és més probable que el L-H.

En aquest cas potser també valdria la pena canviar el que havíem suposat anteriorment. Havíem dit que l'avantpassat devia tenir una F perquè aquest aminoàcid era compartit per la tonyina i el colom, els quals estaven bastant allunyats evolutivament. Potser el millor seria canviar aquesta hipòtesi i dir que l'avantpassat tenia una L, ja que si la L estava codificada per CUC, podem obtenir una H (CAC) i una F (UUC) amb una sola mutació puntual. I, tot i que abans només necessitàvem marcar tres punts de mutació en l'arbre filogenètic i ara en necessitem quatre, en realitat es necessiten menys mutacions per produir-se els canvis del segon cas.

4. Negre. Aquesta columna és la menys problemàtica; l'aminoàcid coincideix a cada espècie, així que no ens hem de preocupar de possibles canvis.
5. Blau. Aquí tenim A, T, S i K, i el pont entre elles és la T: podem fer el canvi A-T-S (GCU-ACU-UCU, GCC-ACC-UCC, GCA-ACA-UCA i GCG-ACG-UCG) i el canvi T-K (ACA-AAA i ACG-AAG).

Per aquest motiu, potser el millor seria canviar la hipòtesi inicial i dir que l'avantpassat comú tenia una T i no una A, ja que si la T estava codificada per ACA o AAG, podem obtenir tots els altres aminoàcids. Tot i així, si mirem les mutacions que ens estalviariem, ens surt zero: necessitem el mateix nombre de mutacions si mantenim la nostra primera suposició com si la canviem per la segona, tal i com es pot veure en aquests dos arbres filogenètics, on cada número representa el número de mutacions necessàries per a passar de l'aminoàcid de l'avantpassat a la d'una espècie:

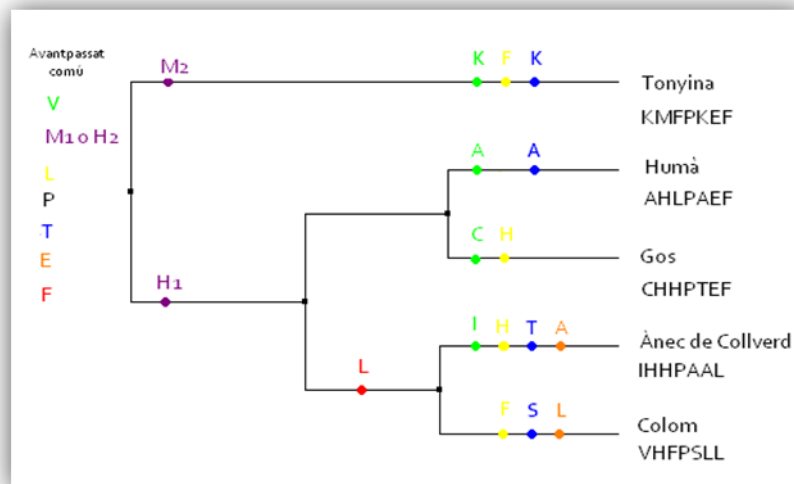


Com que es necessiten el mateix de mutacions, costa dir quin és el més probable que passi evolutivament. En defensa del primer, podem argumentar que és el bo perquè trobem dues A en espècies actuals, mentre que pel segon podem dir que és més important que a partir de la T puguem obtenir amb una sola mutació qualsevol dels aminoàcids que tenen les diferents espècies al mateix punt. Personalment, a mi em convenç més el segon argument, i és per això que l'arbre ha estat modificat.

6. Taronja. Hi trobem E, A i L. El canvi E-A és possible amb una sola mutació puntual (GAA-GCA i GAG-GCG), mentre que el canvi E-L i el L-A en necessiten dues. Pel que teníem pensat, podríem no canviar res perquè es veu clar que la E ha de ser l'aminoàcid de l'avantpassat comú. Ara bé, podríem pensar en un altre tipus de canvi.

Tot i que per fer el canvi E-L i A-L es necessiten en els dos casos dues mutacions, el canvi A-L és més probable, ja que, com es pot veure a la taula del codi genètic, l'àcid glutàmic (E), només és codificat per dos codons, mentre que l'alanina (A) ho és per quatre. Així, podríem pensar que a la branca mare dels ocells va passar una mutació que va canviar l'E per la A, mutació que l'ànec de collverd va conservar però que llavors el colom va canviar per una L. Tot i així, pensar això seria un error, ja que l'A només podria ser codificada per GCA i GCG, però no pels altres dos possibles codons, ja que prové d'una mutació puntual de la E, la qual només té dos codons, com hem dit. Així, seria igual de probable el pas d'E a L que d'A (provenint d'E) a L.

7. Vermell. Aquí hi trobem F i L. La possibilitat d'aquest canvi ja l'hem tractada abans a la columna 3. El canvi F-L pot ser provocat per aquests canvis: UUA o UUG per UUU o UUC i a l'inrevés, o bé CUU-UUU i CUC-UUC, que fan un total de sis possibles canvis. D'aquí que fàcilment s'expliqui aquest canvi d'aminoàcid. En aquest cas no canviarem el que teníem dit, ja que continua sent clar que l'avantpassat tenia una F i que la branca dels ocells va patir una mutació.



## Comentaris

Un cop realitzada, utilitzant un altre mètode, la col·locació de les mutacions en l'arbre filogenètic, es pot destacar que el fet que consultar el codi genètic ha ajudat a fer molt més acurat l'arbre filogenètic, fent necessaris certs canvis o fent creïbles certs canvis d'aminoàcids, fins i tot en casos de gran varietat d'aminoàcids en certes columnes.

Realitzant aquest estudi hem pogut conèixer on s'havien produït els canvis que feien diferenciar les espècies del seu avantpassat comú, però també hem pogut saber més coses de l'altre cara de la moneda: de l'avantpassat comú. D'aquest n'hem pogut arribar a endevinar la seqüència de la zona que correspon a la zona menys conservada de les cinc espècies, tot i que hem obtingut un resultat per a cada mètode:

	1	2	3	4	5	6	7
1 <sup>er</sup> mètode	K	M o H	F	P	A	E	F
2 <sup>on</sup> mètode	V	M o H	L	P	T	E	F

Veiem que els dos mètodes difereixen en les columnes 1, 3 i 5. Com ja hem dit anteriorment, el segon mètode és més acurat i, per tant, s'ha de tenir en compte el resultat d'aquest. A més a més, amb aquest mètode també hem pogut arribar a endevinar part de la seqüència d'ADN de l'avantpassat només treballant amb aminoàcids d'espècies actuals, ja que segurament la Valina d'aquesta seqüència estava codificada per GUU i la Leusina per CUC.

D'aquests resultats també podem mirar la seqüència de quina espècie s'ha conservat millor i quina ha patit més mutacions alineant la seqüència d'aminoàcids de totes les espècies i comparant-les amb els aminoàcids obtinguts amb el segon mètode. Enlloc de dibuixar la taula següent també haguéssim pogut utilitzar l'arbre filogenètic del segon mètode i comptar tots els canvis d'aminoàcid que tenen lloc des de l'avantpassat comú fins a arribar a l'espècie i haguéssim obtingut els mateixos resultats, però la taula ens permet comparar més fàcilment aminoàcid per aminoàcid:

	1	2	3	4	5	6	7
Avantpassat comú	<b>V</b>	<b>M o H</b>	<b>L</b>	<b>P</b>	<b>T</b>	<b>E</b>	<b>F</b>
Humà	A	H	<b>L</b>	<b>P</b>	A	<b>E</b>	<b>F</b>
Gos	C	H	H	<b>P</b>	<b>T</b>	<b>E</b>	<b>F</b>
Ànec de collverd	I	H	H	<b>P</b>	A	A	L
Colom	<b>V</b>	H	F	<b>P</b>	S	L	L
Tonyina	K	M	F	<b>P</b>	K	<b>E</b>	<b>F</b>

En verd trobem els aminoàcids que s'han conservat, deixant de banda la segona columna per la incertesa que trobem amb la decisió de quin aminoàcid tenia l'avantpassat. Pel que sembla, el títol de tenir la seqüència d'aminoàcids més conservada el comparteixen l'hemoglobina humana i la de gos, amb quatre aminoàcids conservats sobre set. I el títol de tenir la seqüència menys conservada se l'emporta l'ànec de collverd, amb només un aminoàcid conservat, que és l'aminoàcid que totes les seqüències tenien igual.

S'ha de remarcar que totes aquestes conclusions parteixen dels resultats del segon mètode que, tot i basar-se en el codi genètic, continuen sent suposicions i, per tant, no es pot assegurar que siguin els reals.

## Situació de les zones perfectament conservades a l'estructura terciària

L'objectiu d'aquest apartat és intentar donar una explicació a la conservació de quatre zones de cada cadena A de totes les hemoglobina. Per tant, el que farem és situar les zones perfectament conservades a l'estructura terciària de la cadena A de cada hemoglobina utilitzant el programa VMD.

Com es podia veure a l'apartat de "zones estudiades", les zones perfectament conservades (marcades en verd) es consideraven d'aquest grup a partir de tenir de quatre a cinc aminoàcids següents idèntics en totes les espècies. A més, es podia comprovar que estaven bastant disperses per la proteïna.

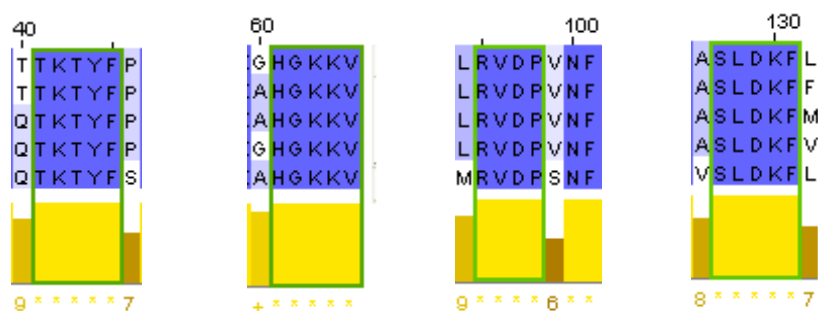
La primera zona a estudiar se situa als aminoàcids 39-43 d'ocells i mamífers i als 41-45 de la tonyina. Per tant, comprèn cinc aminoàcids. En les imatges la trobarem pintada de vermell.

La segona zona es troba als aminoàcids 58-62 d'ocells i mamífers i als 61-65 de la tonyina. També comprèn cinc aminoàcids. En les imatges està pintada de verd.

La tercera zona comprèn els aminoàcids 92-95 d'ocells i mamífers i els 95-98 de la tonyina. En aquest cas es tracta de quatre aminoàcids, formant la zona perfectament conservada més petita. En les imatges la veurem de lila.

La quarta i última zona es troba situada als aminoàcids 124-128 d'ocells i mamífers i als 127-131 de la tonyina. En les imatges estarà destacada amb blau.

Els aminoàcids aquí nombrats són els extrems de l'alineament fet pel JalView. A l'hora ressaltar els aminoàcids amb el VMD, hem de recordar que aquest programa no té en compte l'aminoàcid indefinit de la tonyina. Per tant, hem de recordar de restar un aminoàcid en totes les zones de la tonyina.



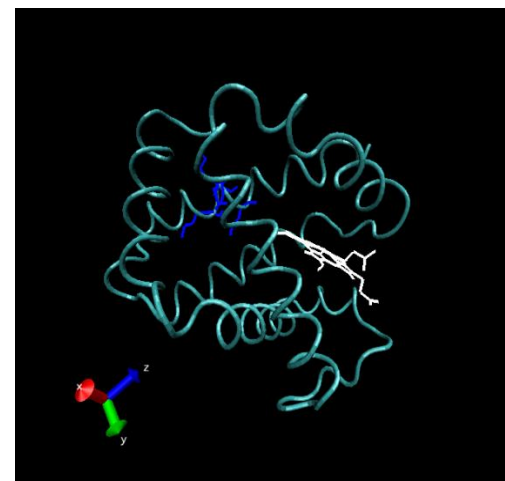
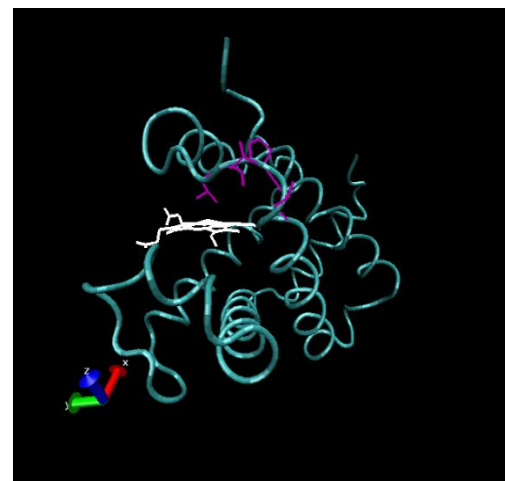
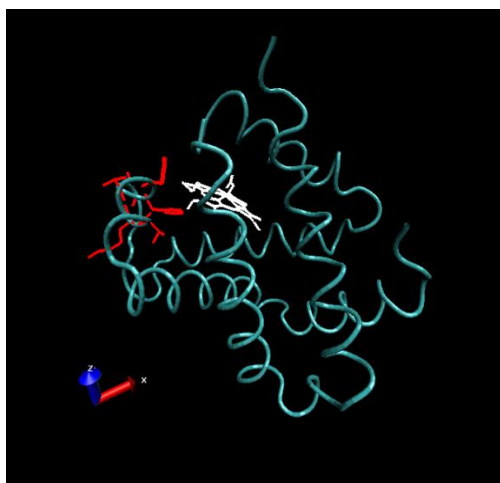
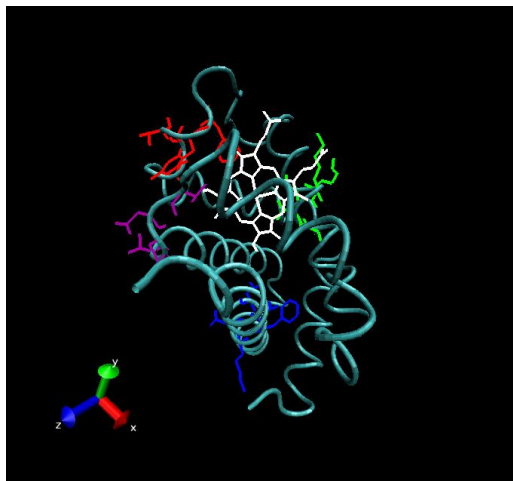
### Resultats esperats

La nostra hipòtesi és que les zones perfectament conservades han de ser zones de la part catalítica de la proteïna, i no de l'estructural. Un canvi en aquestes zones podria suposar que el grup hemo no pogués tenir la forma correcta i necessària per a la fixació de oxigen, fet que provocaria que l'individu morís i que, per tant, la mutació no passés a la descendència. Tot i que és veritat que una mutació en la part estructural de la proteïna pot acabar afectant la part catalítica, creiem que, si la zona és tan altament conservada, és perquè ha d'estar relacionada directament amb la funció de la proteïna. La manera de comprovar si formen la part catalítica serà veure la seva situació en l'estructura terciària; segons el que esperem, les zones han d'estar situades molt a prop del grup hemo.

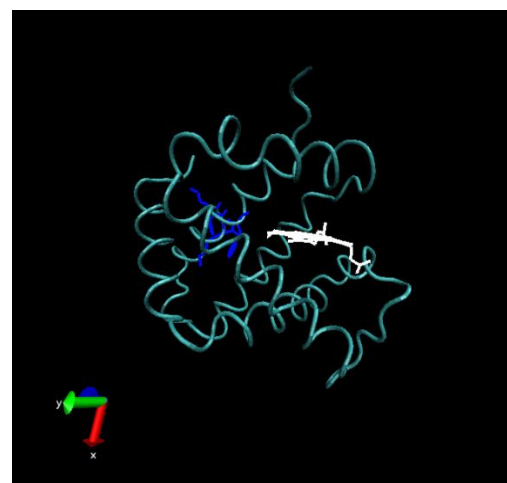
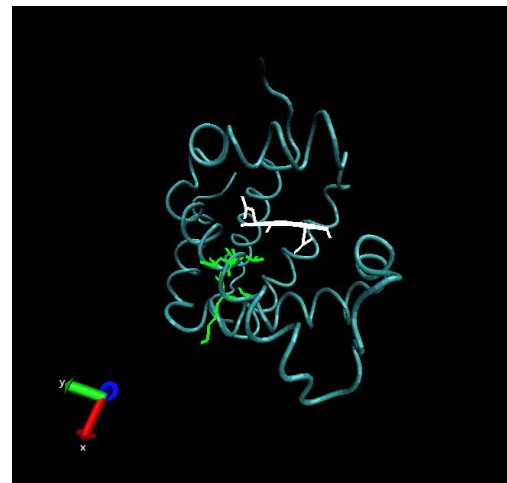
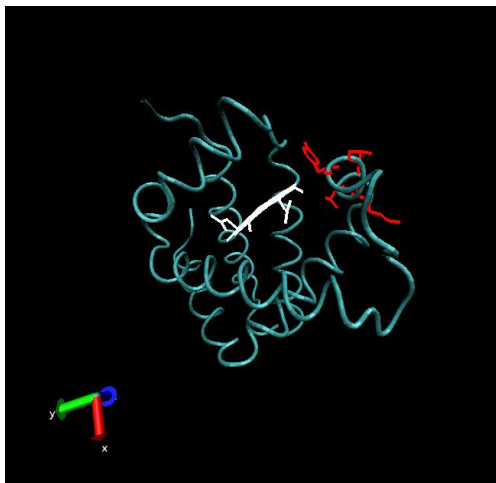
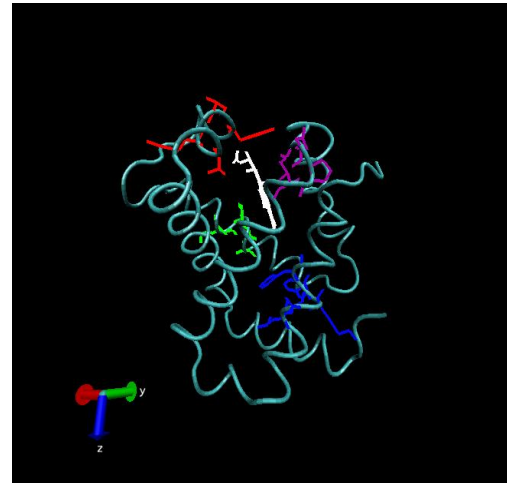
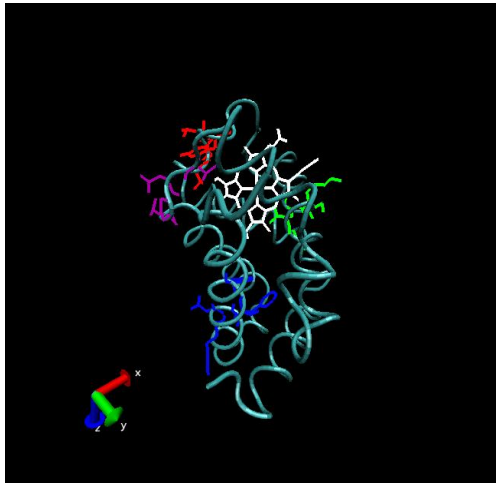
### Resultats obtinguts

En aquest cas també hem guardat sis imatges de cada proteïna. En les dues primeres es veuen dues perspectives de la proteïna amb les quatre zones d'interès marcades. Les quatre següents són la proteïna cada vegada una zona diferent destacada.

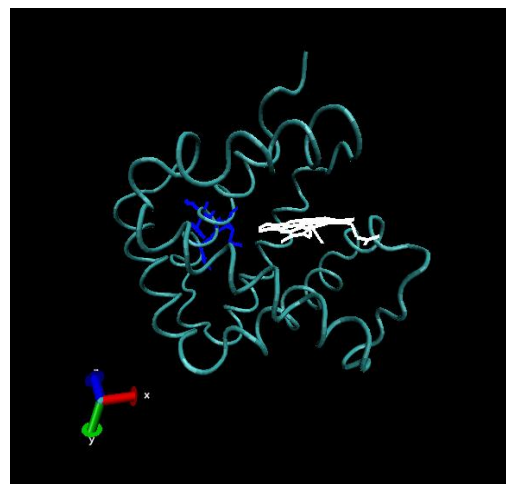
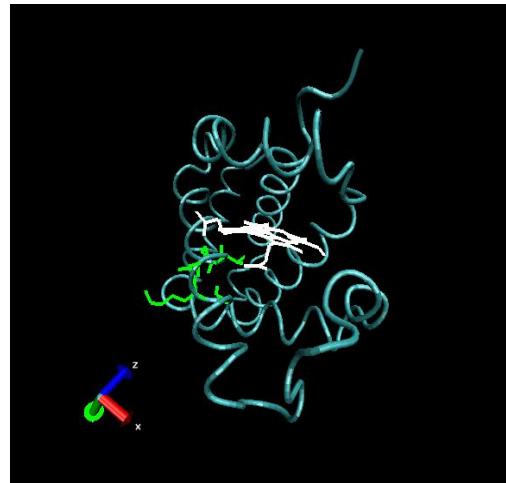
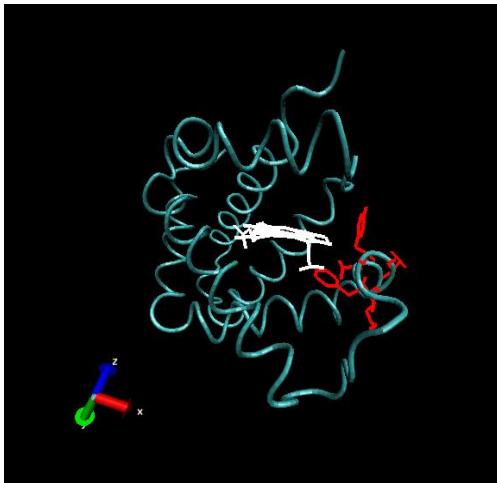
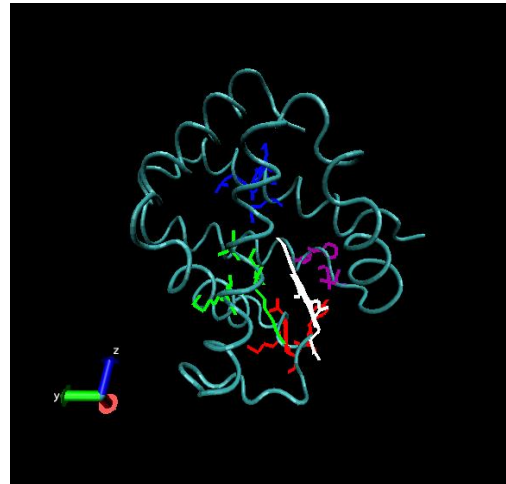
#### ***Humà***



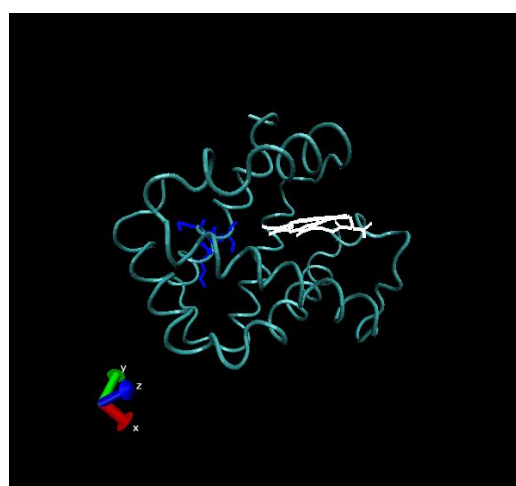
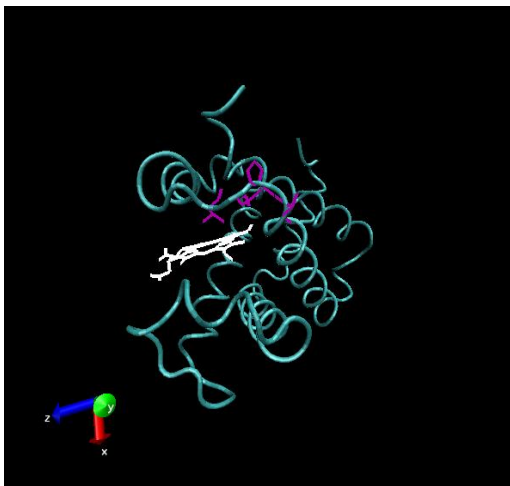
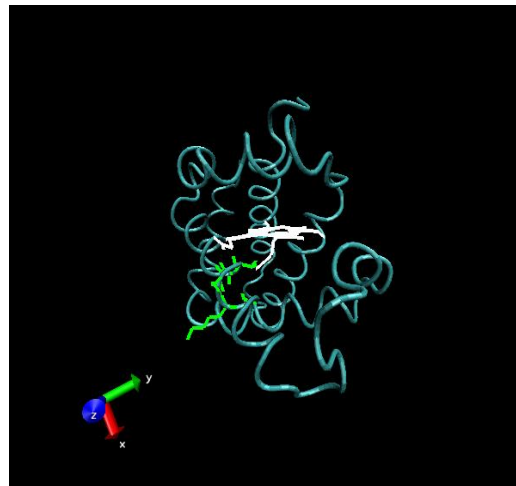
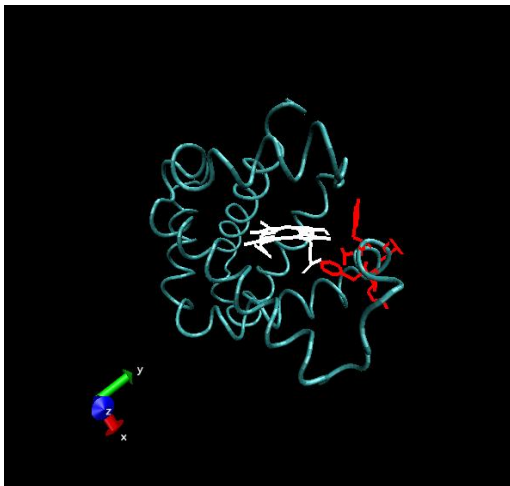
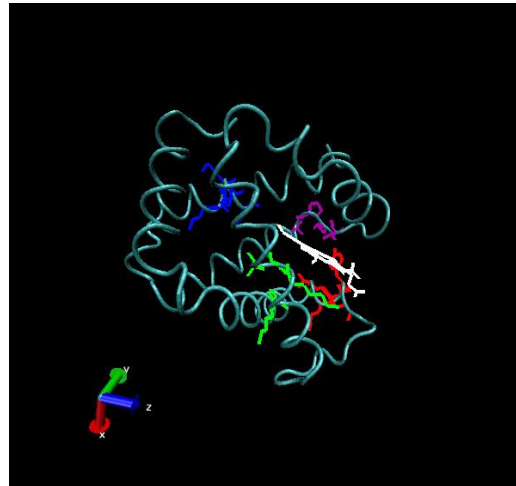
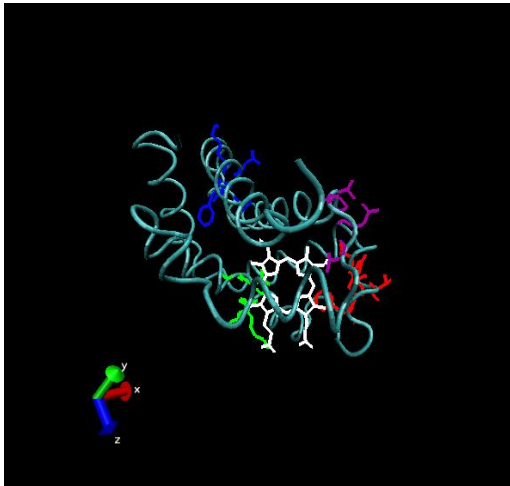
*Gos*



*Ànec de collverd*

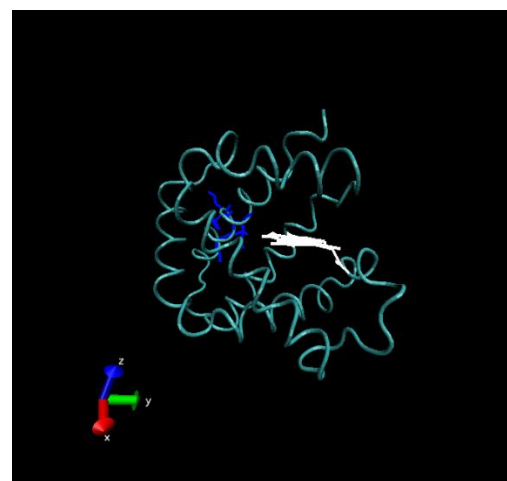
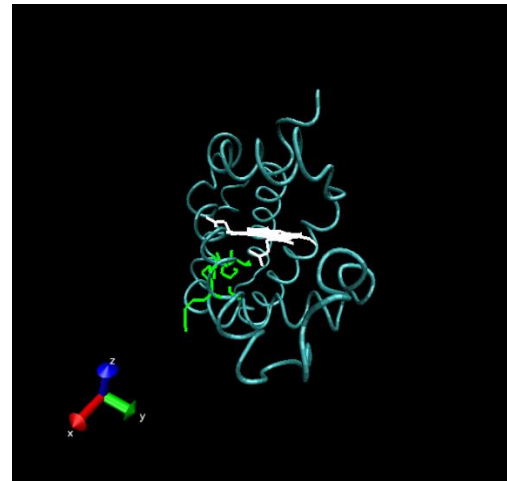
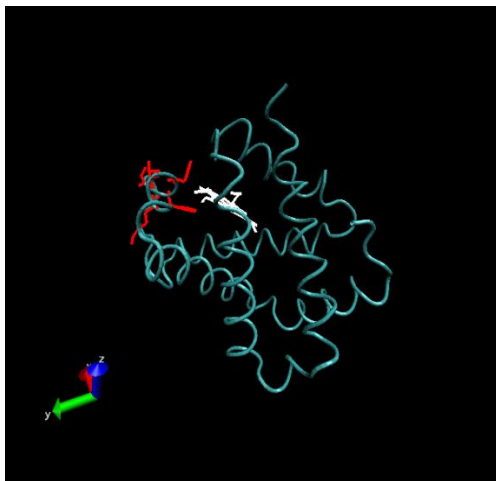
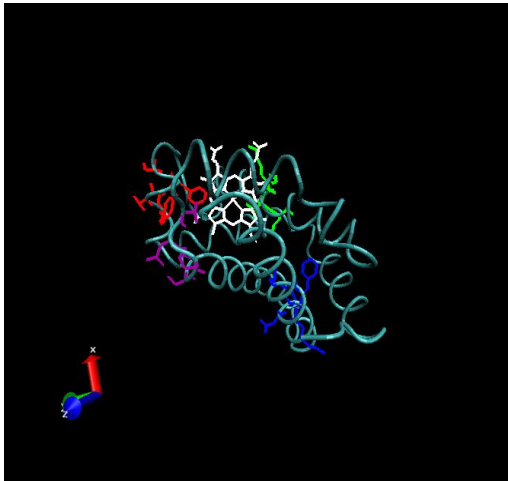


*Colom*





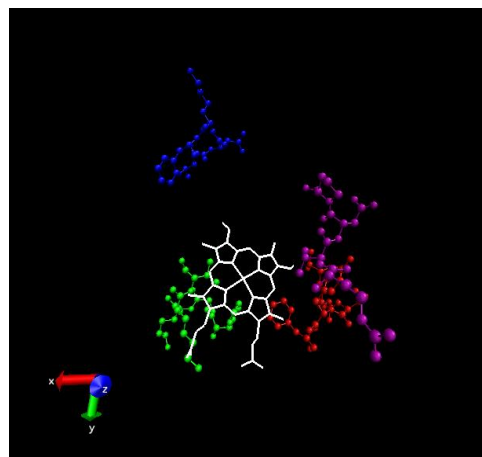
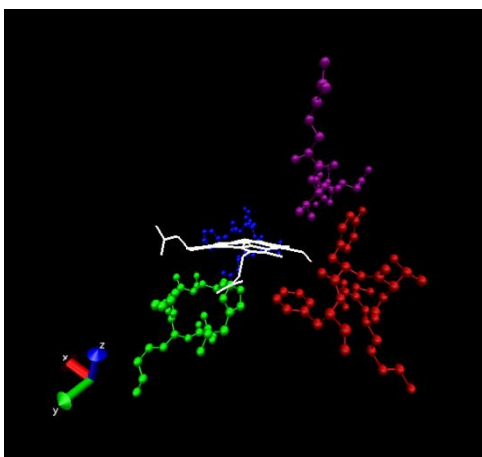
**Tonyina**



## Comentaris

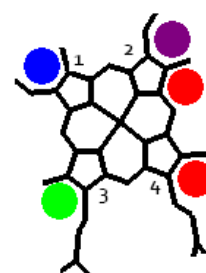
Confirmant la nostra hipòtesi, podem veure indubtablement que les quatre zones perfectament conservades es troben a les hèlix  $\alpha$  més pròximes al grup hemo. També hem vist que la posició d'una zona concreta era la mateixa en totes les espècies.

Per tal de veure més clar la posició de les zones respecte el grup hemo, aquí sota hi ha dues perspectives de la proteïna humana on només hi ha ressaltats el grup hemo i les zones d'interès. D'aquesta manera, si eliminem la resta de molècula podem veure molt fàcilment la posició de les zones. Aquí, enlloc de veure els aminoàcids representats per línies, veiem que cada aminoàcid és una boleta. Tot i que aquesta imatge sigui de l'hemoglobina humana, també ens servirà per veure la posició de les zones de les altres espècies perquè, com hem vist en les imatges anteriors, la posició d'aquestes no variava entre espècies. Així, veiem clarament en la primera imatge que la quarta zona (color blau) és l'única que es troba en línia amb el grup hemo. La primera (en vermell) i la segona (en verd) zones es troben per sota del grup hemo, al contrari de la tercera (en lila), que es troba per sobre d'aquest.



Després d'observar la segona imatge i també totes les altres, es pot veure que cada zona perfectament conservada està situada prop d'un extrem del grup hemo, tal i com es pot veure en el dibuix del costat. Els números dels extrems els he decidit fent de d'esquerra a dreta i de dalt a baix. D'aquesta manera, en l'extrem 1 es troba situada la quarta zona; al 2, la primera i la tercera; al 3, la segona i, al 4, la primera. Per tant, totes les zones es troben a un extrem menys la primera, que avarca dos extrems.

D'aquí podem treure que la part més important de l'hemoglobina siguin les zones que es troben prop dels extrems del grup hemo, ja que deuen permetre la bona posició d'aquest dins la proteïna i, per tant, la futura bona fixació de l'oxigen en el Fe del grup hemo.



Imatge 5: Grup hemo amb les zones perfectament conservades (adaptada de 8)

## Flexibilitat de les diferents hemoglobines

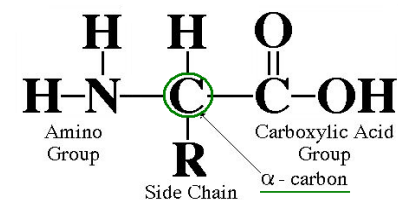
### Introducció al FlexServ

Les proteïnes no només evolucionen per tal aconseguir la seva estructura òptima, sinó que també evolucionen per obtenir propietats de deformació òptimes per a la seva funció. D'aquesta manera, moltes evidències mostren que l'evolució conserva cuidadosament els models de deformació com a mecanismes per mantenir la funció. Això vol dir que si són flexibles són més immunes a certs canvis? Potser, d'aquesta manera, es poden adaptar millor.

Tot i que hi ha hagut avanços en tècniques experimentals, l'estudi de la flexibilitat és una tasca bàsicament de mètodes teòrics. El més important i efectiu és l'MD (*atomistic molecular dynamics*, en anglès), un mètode rigorós amb bases físiques sòlides, el qual permet obtenir representacions molt acurades de la flexibilitat de la proteïna. Per desgràcia, l'MD és complex, car pel que fa als càlculs i es necessita un cert grau d'habilitat per utilitzar-lo. I, com que té en compte tots els àtoms de la molècula, tarda setmanes a obtenir els resultats.

Són per aquests motius que s'han hagut de dissenyar altres mètodes que provoquen assumir una pèrdua de detalls atòmics per tal de guanyar simplicitat en les representacions de la flexibilitat de les proteïnes. Aquests mètodes s'anomenen, en anglès, *coarse-grained methods*.

Una pàgina d'Internet que permet disposar de diversos mètodes *coarse-grained* és el FlexServ, que a més recull informació estructural i models atòmics amb un sistema d'anàlisi poderós. El FlexServ ofereix tres d'aquests mètodes, els quals tenen en comú que, en lloc de considerar tots els àtoms de la proteïna, només tenen en compte el carboni  $\alpha$  ( $C\alpha$ ) de cada aminoàcid, facilitant molt més els càlculs:



Imatge 6: Estructura d'un aminoàcid  
(Font: <http://philomath.exteen.com>)

1. DMD (Discrete dynamics, en anglès): aquest mètode considera que només hi ha interacció entre els  $C\alpha$  veïns i els mateixos  $C\alpha$ . Llavors, els  $C\alpha$  de les diferents hèlix  $\alpha$  van xocant i d'aquí s'extreu la flexibilitat de la proteïna.
2. NMA (Normal mode analysis, en anglès): en aquest sistema, tots els  $C\alpha$  estan units amb enllaços semblants a molles amb els  $C\alpha$  pròxims.
3. BD (Brownian dynamics, en anglès): aquest és el mètode que nosaltres vam utilitzar per a realitzar els nostres càlculs. El BD considera que les proteïnes estan en un bany tèrmic, envoltades d'aigua, a T constant. Per tal de saber el moviment de les partícules utilitza la primera llei de Newton ( $F=m \cdot a$ ), però si només utilitzem aquesta fórmula, la temperatura resultant no seria constant. Llavors, el que fa l'BD és tenir en compte dos factors més:  $F - \gamma \cdot v + \zeta = m \cdot a$ . La  $\gamma$  (gamma) multiplicada per la velocitat és el terme de fricció amb l'aigua, i la  $\zeta$  (zeta) és un número aleatori que representa els xocs aleatoris de les molècules d'aigua a la proteïna.

Com ja hem dit abans, tots aquests mètodes es basen en la física, i el meu nivell en aquest camp no era el suficient per a poder entendre exactament com aquests obtenen els resultats. Tot i així, l'objectiu no era aquest sinó ser capaç de comprendre els resultats i saber-ne treure conclusions. Per tant, no em vaig endinsar en la part teòrica dels diferents mètodes.

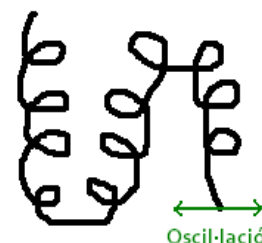
## Factors estudiats pel FlexServ

FlexServ estudia diversos factors de flexibilitat de la proteïna, bàsicament els que tenen a veure amb els punts amb més moviment de la proteïna i les interaccions entre aminoàcids. En aquest apartat explicaré els factors que vaig utilitzar de l'estudi del FlexServ, ja que no tots m'eren útils: el *B-factors landscape*, *Lindemann coefficients*, *Apparent stiffness* i *Residue correlations*.

Tot i que el FlexServ només té en compte els C $\alpha$  de cada aminoàcid, en els resultats parlarem de l'"aminoàcid n<sup>o</sup>X" i no de "el C $\alpha$  de l'aminoàcid n<sup>o</sup>X".

### B-factors landscape

El B-factor és el nivell de mesura de flexibilitat d'un àtom o residu, en el nostre cas, un C $\alpha$ . Es determina a partir de les oscil·lacions del C $\alpha$  respecte la seva posició d'equilibri; l'amplitud de l'oscil·lació. Com més gran és el valor del B-factor, més es mou el C $\alpha$ . Aquest factor es representa amb una gràfica, on les unitats són Àngstroms al quadrat (Å<sup>2</sup>).



Una altra manera de veure reflectit el B-factor és amb el *Backbone coloured according to the B-factors*. Aquí veiem tota la seqüència de la proteïna, amb els aminoàcids pintats segons el seu B-factor: els colors van de blanc (gens de moviment) a vermell (molt de moviment). A més, s'acompanya la seqüència amb una representació de Java del moviment de la molècula, amb els aminoàcids pintats de la mateixa manera.

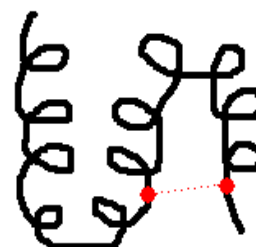
### Lindemann coefficients

Aquest coeficient és una estimació del comportament sòlid-líquid de la proteïna, de manera que ens indica la rigidesa de la proteïna. Valors per sota de 0.15 es consideren sòlids i són pintats de marró, mentre que els valors per sobre de 0.15 són els líquids i es pinten de blau. Per tant, com més alt és el valor menys rígida és la molècula. Que la molècula sigui poc rígida és bo perquè és més flexible i, d'aquesta manera, més funcional.

Si s'estudien els residus de la proteïna es pot veure que els residus de l'interior de la proteïna tenen el coeficient més baix i els de l'exterior més alt, cosa que és fàcil d'explicar perquè la densitat de la proteïna a l'interior és més alta i, per tant, menys flexible, al contrari dels aminoàcids de la superfície. Com que l'hemoglobina no té un nucli diferenciat sinó que són tot d'hèlix  $\alpha$  exteriors, no tindria sentit fixar-se amb aquest contrast d'interior-exterior.

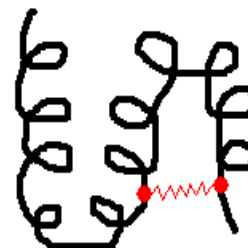
### Apparent stiffness

Aquest factor representa la força promig que actua entre, en el nostre cas, dos C $\alpha$ . Per representar-ho es dibuixa una taula on en els dos eixos hi ha tots els C $\alpha$ , de manera que es pot veure la interacció entre dos C $\alpha$  qualssevol. Els colors es mouen de blanc a vermell: blanc és igual a interacció nul·la i vermell a interacció molt forta. A més a més, podem saber exactament el valor de la força i els dos C $\alpha$  en qüestió situant-nos damunt del punt d'interès de la representació amb el ratolí.



### Residue correlations

Aquest factor també es representa en una taula amb els C $\alpha$  en els dos eixos, com l'*Apparent Stiffness*. A diferència d'aquest, que representava la força entre els C $\alpha$ , el *Residue correlation* mostra les connexions entre els moviments dels C $\alpha$ . Així, podem veure per quins parells de C $\alpha$ , el moviment d'un dels dos afecta més a l'altre. En la representació hi trobarem colors de blanc a verd i vermell: el blanc indica que els dos C $\alpha$  no estan gens correlacionats i que, per tant, si un dels dos es mou, no afectarà per res a l'altre, mentre que el verd i vermell indiquen una correlació forta. Que sigui verd o vermell depèn de si la correlació és positiva (verd) o negativa (vermell), ja que depèn d'una qüestió de correlació o anticorrelació, però això no ens és d'interès. En el web també podem saber exactament el valor de la correlació entre un parell de C $\alpha$  concrets situant-nos damunt del punt d'interès de la representació amb el ratolí.



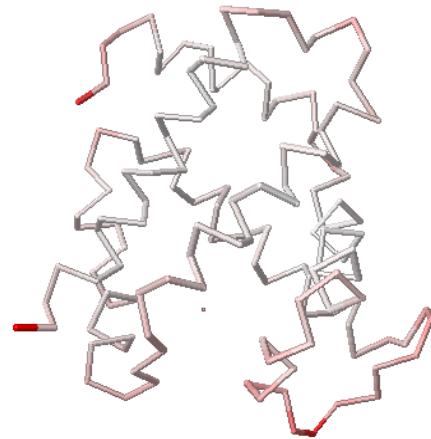
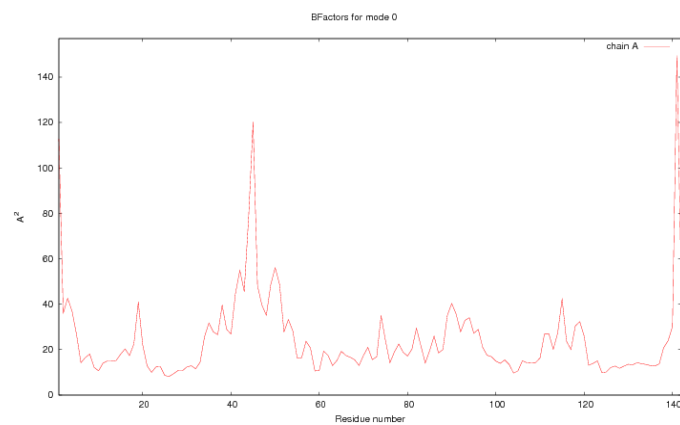
### Estudi del grup hemo

Com ja hem dit, el BD només té en compte els C $\alpha$ . D'aquesta manera, tots aquests factors només tindran en compte la seqüència d'aminoàcids, deixant de banda el grup hemo, el qual no detectarà perquè no té cap C $\alpha$ . Tot i així, també és interessant veure si el grup hemo es mou i, sobretot, amb quins aminoàcids té més interacció i correlació. Per tant, el que vam fer va ser modificar el fitxer pdb de la cadena A de cada hemoglobina, i vam substituir tot el grup hemo per C $\alpha$  que vam situar en la posició del Fe per tal que estigués en la part que correspondria al centre del grup hemo. Aquest C $\alpha$  fa la funció d'àtom estructural, ja que permet conservar una mica part de l'estructura i, el més important, és detectat pel FlexServ.

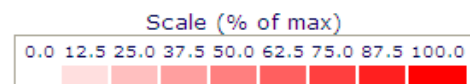
Per tal de veure representat aquest C $\alpha$  en tots els factors anteriorment citats, vam afegir al fitxer modificat que es tractaria d'una Glicina, com podria ser qualsevol altre aminoàcid. Així, en tots els factors trobarem que l'últim aminoàcid és una Glicina. En realitat, serà aquest C $\alpha$  que representa el grup hemo. Com els mateixos resultats del BD, la substitució del grup hemo per un C $\alpha$  no és més que una aproximació molt forta, ja que de ben segur el grup hemo, amb la seva mida superior a un aminoàcid i amb les seves propietats químiques, difereix d'un carboni. No obstant això, aquesta era l'única manera de poder introduir un representant del grup hemo en els resultats del FlexServ.

## *B-factors landscape*

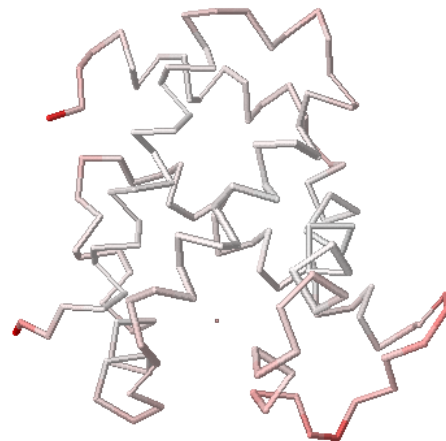
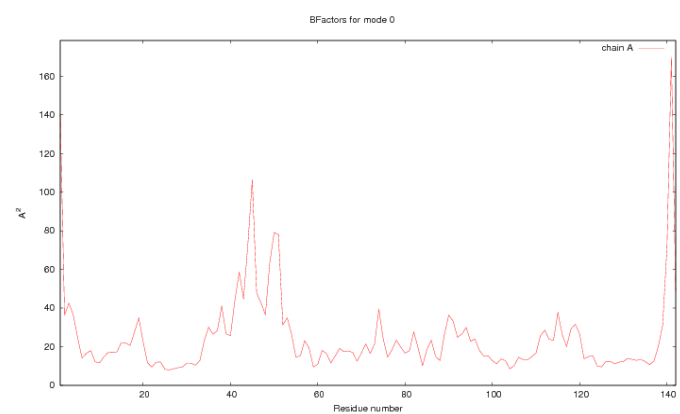
### Humà



A1	V	L	S	P	A	D	K	T	N	V	K	A	A	W	G	K	V	G	A	H	A	G	E	Y	G	A	E	A	L	E	R	M	F	L	S	F	P	T	T	K
A41	T	Y	F	P	H	F	D	L	S	H	G	S	A	Q	V	K	G	H	G	K	V	A	D	A	L	T	N	A	V	A	H	V	D	D	M	P	N	A	L	
A81	S	A	L	S	D	L	H	A	H	K	L	R	V	D	P	V	N	F	K	L	L	S	H	C	L	L	V	T	L	A	A	H	L	P	A	E	F	T	P	A
A121	V	H	A	S	L	D	K	F	L	A	S	V	S	T	V	L	T	S	K	Y	R	G																		

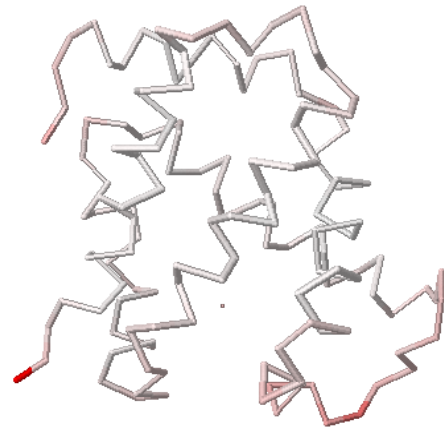
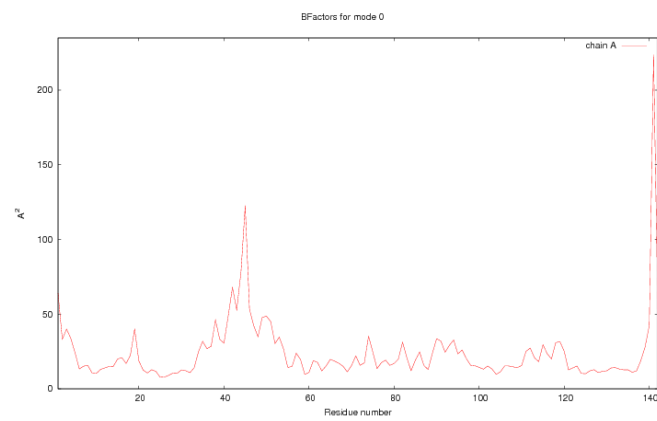


### Gos



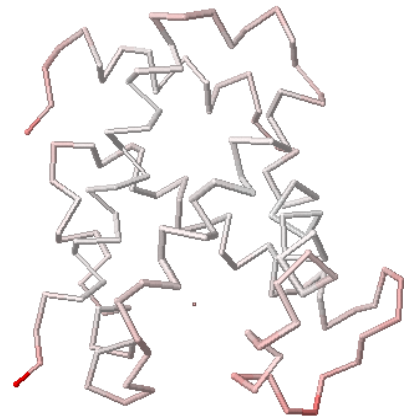
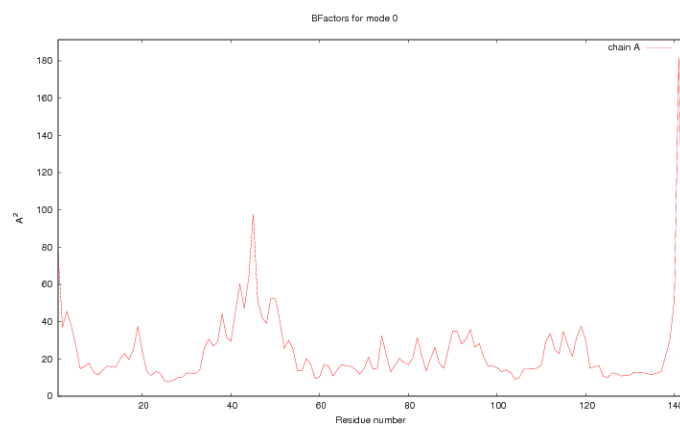
A1	V	L	S	P	A	D	K	T	N	I	K	S	T	W	D	K	I	G	G	H	A	G	D	Y	G	G	E	A	L	D	R	T	F	Q	S	F	P	T	T	K
A41	T	Y	F	P	H	F	D	L	S	P	G	S	A	Q	V	K	A	H	G	K	K	V	A	D	A	L	T	T	A	V	A	H	L	D	D	L	P	G	A	L
A81	S	A	L	S	D	L	H	A	Y	K	L	R	V	D	P	V	N	F	K	L	L	S	H	C	L	L	V	T	L	A	C	H	H	P	T	E	F	T	P	A
A121	V	H	A	S	L	D	K	F	F	A	A	V	S	T	V	L	T	S	K	Y	R	G																		

## Ànec de collverd



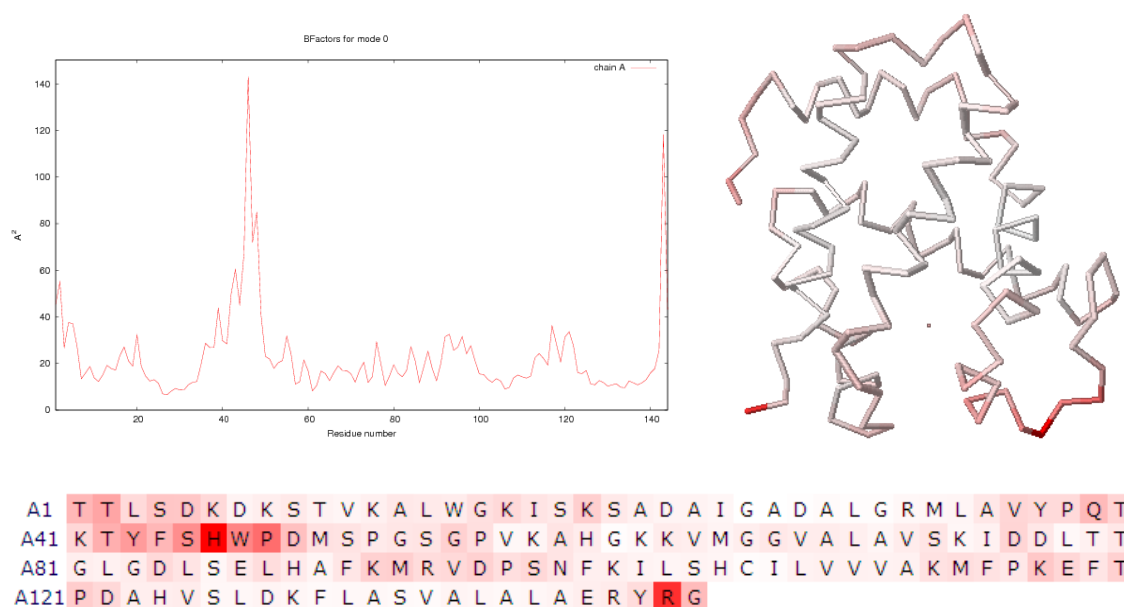
A1 V L S A A D K T N V K G V F S K I G G H A E E Y G A E T L E R M F I A Y P Q T K  
A41 T Y F P H F D L S H G S A Q I K A H G K K V A A A L V E A V N H V D D I A G A L  
A81 S K L S D L H A Q K L R V D P V N F K F L G H C F L V V V A I H H P A A L T P E  
A121 V H A S L D K F M C A V G A V L T A K Y R G

## Colom



A1 V L S A N D K S N V K A V F A K I G G Q A G D L G G E A L E R L F I T Y P Q T K  
A41 T Y F P H F D L S H G S A Q I K G H G K K V A E A L V E A A N H I D D I A G A L  
A81 S K L S D L H A Q K L R V D P V N F K L L G H C F L V V V A V H F P S L L T P E  
A121 V H A S L D K F V L A V G T V L T A K Y R G

## Tonyina



## Comentaris

El que podem veure per a cada espècie és primer el gràfic del B-factor a l'esquerra, al seu costat es troba la imatge congelada del vídeo del moviment de la proteïna pintada segons el B-factor i, a sota, la seqüència d'aminoàcids pintada igualment segons el B-factor. La glicina, l'aminoàcid que representa el grup hemo, es troba al final del gràfic i de la seqüència i, en la imatge congelada, es pot distingir un puntet una mica avall del centre de la imatge: la glicina. Podem trobar l'escala de color sota la seqüència d'aminoàcids de l'humà.

A l'hora de comparar les gràfiques del B-factor, s'ha de tenir en compte que els valors de l'escala de l'eix Y són diferents en cada cas. Aquest canvi d'escala és causat pel valor de l'arginina, l'últim aminoàcid de la cadena, que quan pren valors molt grans fa necessari un augment dels valors de l'escala.

El que s'ha de tornar a ressaltar és que analitzant la seqüència de la tonyina, es veu que aquesta té, deixant de banda la glicina final, 143 aminoàcids. Això ens indica que, com passa amb el VMD, el FlexServ també deixa de banda l'aminoàcid indefinit de la tonyina.

### **Zones amb B-factor alt**

Si analitzem la gràfica del B-factor veiem que en general hi ha tres zones amb el B-factor alt: als primers aminoàcids, entre els aminoàcids 40 i 60 i als últims aminoàcids. Pel que fa a la primera zona, si mirem més atentament veiem que és el primer aminoàcid el que té la oscil·lació més alta, excepte en el cas de la tonyina, on l'aminoàcid en qüestió és el segon. Aquest fet és fàcil d'explicar, ja que el segon aminoàcid de la tonyina correspon al primer de les altres espècies, recordant que aquí no es té en compte l'aminoàcid incert de la tonyina.



Per saber exactament a quin aminoàcid correspon el pic de la segona zona, el millor és mirar la seqüència d'aminoàcids. Si busquem un aminoàcid entre el 40 i el 60 pintat de vermell fort veurem que, en els quatre casos, correspon a una Histidina. Utilitzant els números que encapçalen cada fila d'aminoàcids, podem determinar que aquesta Histidina correspon a l'aminoàcid 45 per ocells i mamífers i al 46 per a la tonyina.

El punt més alt de la tercera zona sempre correspon al penúltim aminoàcid. De fet, correspon a l'últim aminoàcid de la cadena, una arginina (R), ja que l'últim aminoàcid que veiem a la seqüència és una glicina, l'aminoàcid que representa el grup hemo.

Que el primer i l'últim aminoàcids tinguin altes oscil·lacions és normal, ja que es troben a la punta de la cadena i, per tant, no estan lligats a enlloc. En la imatge congelada, el primer aminoàcid es troba a l'extrem de la cadena de dalt a l'esquerra, mentre que l'últim se situa a l'extrem de baix a l'esquerra. Ara bé, com és que l'His 45/46 es mou tant? A més, podem veure que tots els aminoàcids del voltant d'aquest també tenen oscil·lacions força altes. La resposta es troba a la imatge congelada del vídeo. Allà hi veiem que aquesta zona se situa a la part inferior dreta de la proteïna. El més probable és que, com que el grup hemo només està representat per un carboni, aquesta part quedi menys fixada a l'hemoglobina del que hi quedaria normalment.

### ***Comparació dels B-factors de les diferents proteïnes***

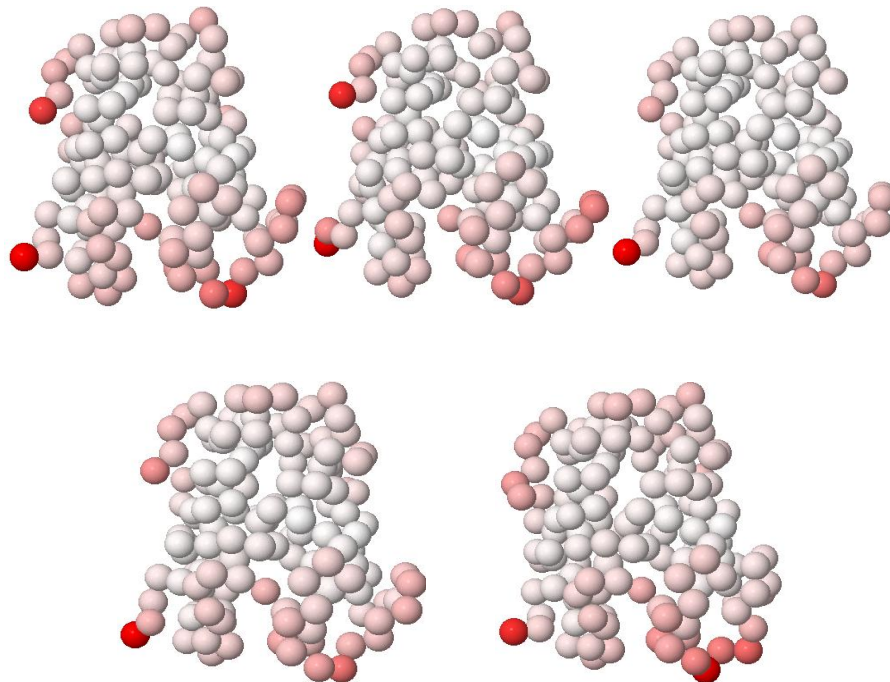
Com dèiem abans, els valors que més difereixen entre espècies són els de l'últim aminoàcid. Tot i així, els valors dels dos altres pics també presenten una gran varietat. Els resultats, d'un valor aproximat, són els següents:

	1er pic (Å <sup>2</sup> )	2on pic (Å <sup>2</sup> )	3er pic (Å <sup>2</sup> )
Home	110	120	150
Gos	140	110	170
Ànec de collverd	60	125	220
Colom	80	100	180
Tonyina	55	140	120

La tonyina té els valors més baixos pel 1<sup>er</sup> i 3<sup>er</sup> pics i, en canvi, té el valor més alt pel 2<sup>on</sup> pic. El colom, amb valors baixos pel 1<sup>er</sup> i 2<sup>on</sup> pic, té un valor considerablement alt pel 3<sup>er</sup>. L'ànec de collverd, amb un valor molt baix pel 1<sup>er</sup> pic, supera amb escreix a totes les espècies en el 3<sup>er</sup> pic. L'home i el gos es mantenen més constants en tots els valors.

El que està clar és que no es pot establir una lògica per dir que un pic sempre tingui el valor més alt, ja que en totes les espècies, l'ordre de petit a gran dels valors dels pics és diferent. Pel que sembla, l'únic que comparteixen les espècies són els aminoàcids que conformen aquests pics. Aquests sí que són sempre els mateixos.

Triant una opció en el FlexServ on cada aminoàcid és representat per una bola (*CPK Spacefill*) podem veure el vídeo del moviment de la proteïna. El que ens permet aquesta aplicació és veure molt clarament la distribució dels colors per la proteïna. Les imatges corresponen, en aquest ordre, a humà, gos, ànec de collverd, colom i tonyina:



Els aminoàcids vermells ja els teníem situats, però ara veiem molt millor que tant l'ànec de collverd com la tonyina tenen molt poc B-factor al 1<sup>er</sup> pic, ja que mentre que els mamífers el tenen d'un vermell potent, l'ànec de collverd i la tonyina el tenen rosat.

També podem distingir que les zones rosades (oscil·lació mitjana) es troben a la part superior i inferior de l'hemoglobina, i que les blanques (molt poca oscil·lació) se situen a la zona central d'aquesta. Tot i que els aminoàcids de la zona perifèrica no tenen tanta llibertat de moviment com els de l'extrem de la cadena, la seva oscil·lació és fàcilment superior que la dels aminoàcids centrals, els quals estan més estrets i mantenen més l'estructura de la proteïna.

Si observem les hemoglobines en general veurem que l'ànec de collverd i el colom són les espècies amb els tons més clars, denotant poca oscil·lació.

Pel que fa al moviment del grup hemo representat per la glicina, podem veure que no està immòbil, sinó que té una oscil·lació mitjana, a jutjar pel color rosat amb el que està pintat en totes les espècies.

## *Lindemann Coefficients*

Humà

Standard (Å)	
All	0.213

Gos

Standard (Å)	
All	0.213

Ànec de collverd

Standard (Å)	
All	0.212

Colom

Standard (Å)	
All	0.212

Tonyina

Standard (Å)	
All	0.209

### Comentaris

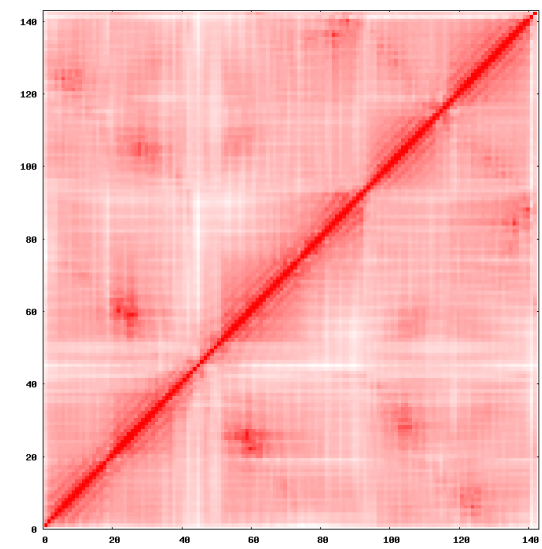
Pel que podem veure, totes les espècies tenen una rigidesa clarament líquida, ja que passen el 0.15. Tot i així, no són valors considerats massa alts. De manera que, tot i ser líquides, no són molt flexibles.

Podem veure que l'hemoglobina d'humans i gossos són les menys rígides amb 0.213, seguides de molt a prop per l'ànec de collverd i el colom amb 0.212 i, finalment, per la de la tonyina amb 0.209. Així, els mamífers tenen el mateix valor i els ocells també. A més, la distància d'aquests dos grups amb la tonyina és superior a la que trobem entre ells dos. D'aquesta manera, tornem a confirmar els resultats de l'arbre filogenètic obtingut amb el ClustalW2.

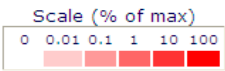
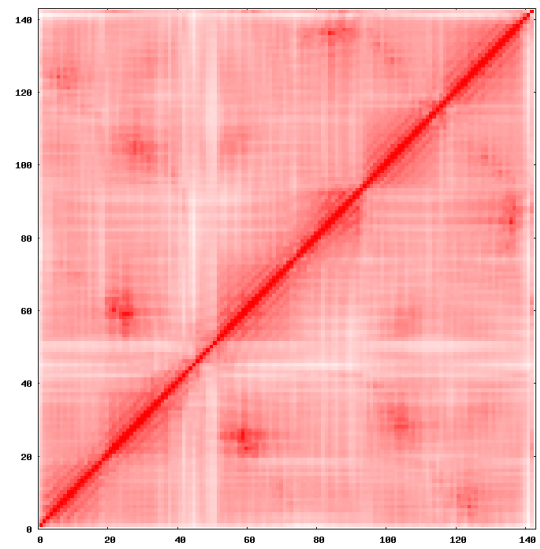
Si ens hi fixem bé, hem dit que les hemoglobines de l'ànec de collverd i del colom eren que tenien un B-factor més baix en general, ja que tenien tons vermellosos molt clars. Així, com s'entén que siguin més flexibles que la de la tonyina? Doncs segurament es tracta dels moviments considerablement superiors als de la tonyina que tenien alguns pics de les hemoglobines dels ocells. Així, aquests grans moviments fan pujar la mitjana de *Lindemann coefficient* que tenen l'ànec de collverd i el colom.

*Apparent Stiffness*

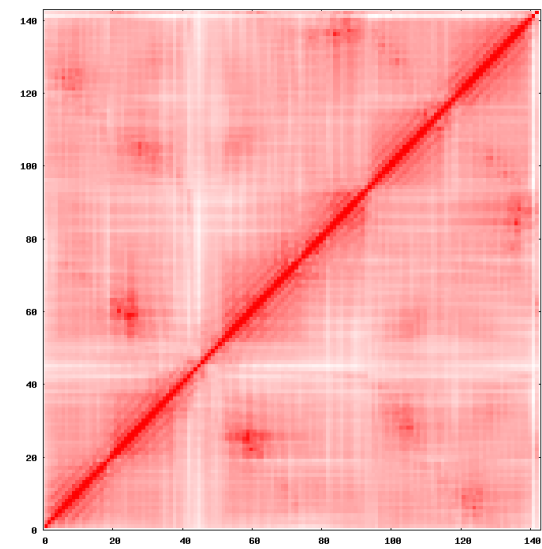
Humà



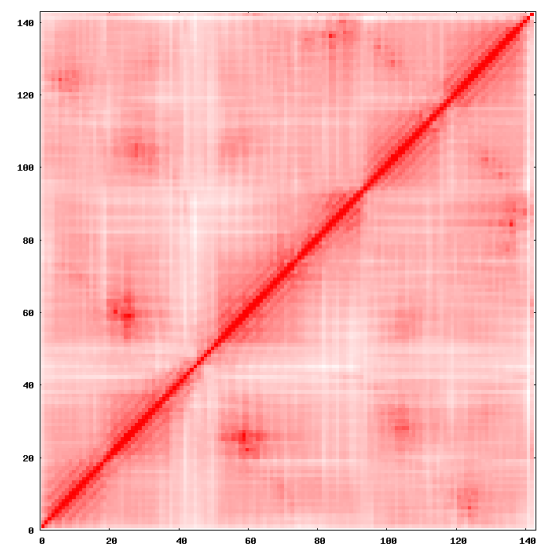
Gos



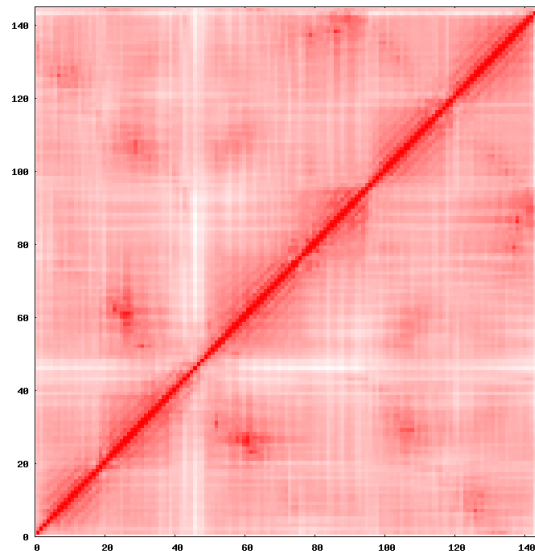
Ànec de collverd



Colom



## Tonyina



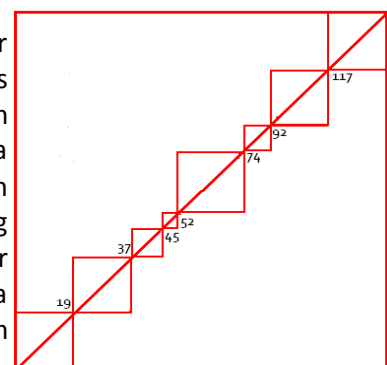
### Comentaris

Com podem veure, en totes les imatges hi ha una línia que divideix la figura en dues parts simètriques. Aquesta simetria és causada pel fet que als dos eixos hi trobem representats els mateixos aminoàcids. D'aquesta manera, és el mateix buscar a l'eix de les X l'aminoàcid 20 i veure la seva interacció amb l'aminoàcid 30 de l'eix de les Y, com buscar l'aminoàcid 30 a l'eix de les X i buscar la seva interacció amb l'aminoàcid 20 de l'eix de les Y. Per això, aquesta simetria neix a partir de la línia central, que és on es troba un aminoàcid amb ell mateix.

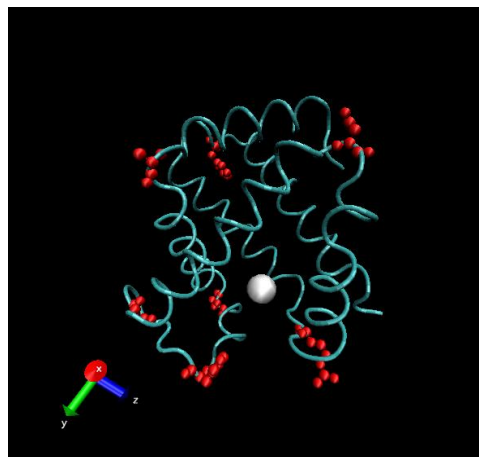
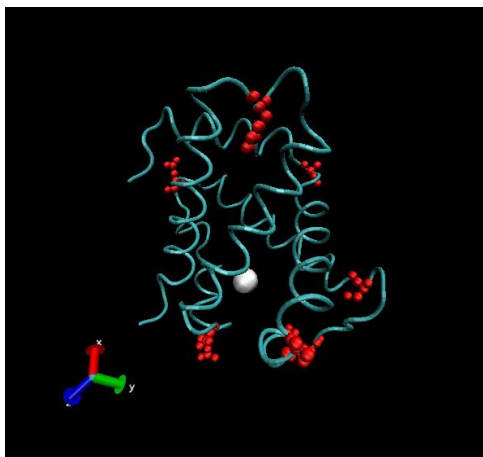
El fet que la línia central sempre sigui d'un vermell tan fort és fàcil d'explicar: la interacció d'un aminoàcid amb ell mateix es considera infinita, i és molt gran amb els aminoàcids adjacents o molt propers. A més, també trobem que és com si aquesta línia central fos la diagonal d'una sèrie de quadrats que van apareixent al llarg de la línia. Per exemple, agafem la gràfica de l'hemoglobina humana i busquem les Ala 19 als dos eixos. Si tracem una línia fins que es trobin, la zona compresa serà un dels quadrats que dèiem. Dins d'un quadrat, tots els aminoàcids tenen una interacció forta amb tots.

### ***Quadrats a l'Apparent Stiffness***

Per saber el per què d'aquests quadrats vam agafar l'hemoglobina humana com a exemple i vam mirar quins aminoàcids aproximadament marcaven el principi o final d'un quadrat, per després buscar la seva situació en l'estructura terciària de l'hemoglobina utilitzant el VMD. Els aminoàcids en qüestió eren l'Ala 19, la Pro 37, l'His 45, la Gly 52, l'Asp 74, l'Arg 92 i la Phe 117, deixant per suposats la Val 1 i l'Arg 141, per ser el primer i l'últim aminoàcid. Com es pot veure, la distància entre dos aminoàcids de vegades és molt curta. També vam intentar trobar-hi l'explicació mirant l'estructura terciària.



Imatge 9: Simplificació dels quadrats que es troben al voltant de la línia central de l'AS (Font: Dibuix realitzat amb Paint)



Aquestes són dues cares obtingudes de l'hemoglobina humana, les quals ens donen una explicació clara d'on és que comença o acaba un quadrat. La circumferència blanca és la glicina que representa el grup hemo, i cada grup d'esferes vermelles representen un aminoàcid, on cada bola és un àtom. Mirant on es troba cada aminoàcid, veiem sense cap mena de dubte que cada un d'ells es troba a una zona on hi ha un canvi d'una hèlix  $\alpha$  a una altra. I la causa dels quadrats més petits és simplement que els dos aminoàcids que els comprenen es troben un al principi i l'altre al final d'una hèlix  $\alpha$  curta, com es pot veure a la segona imatge a la part inferior esquerra de la imatge.

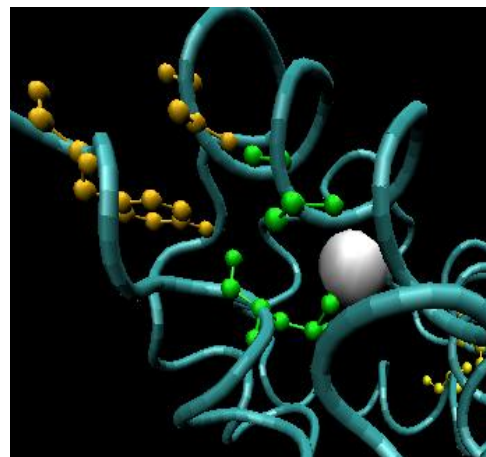
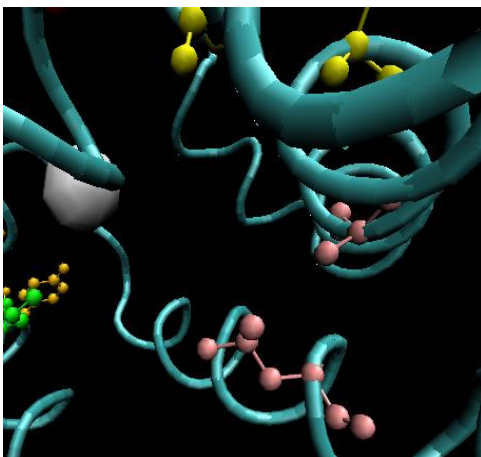
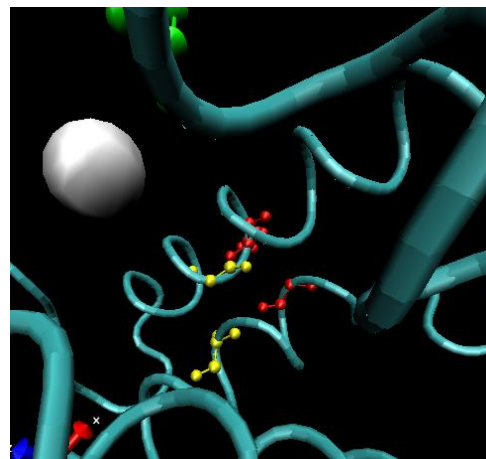
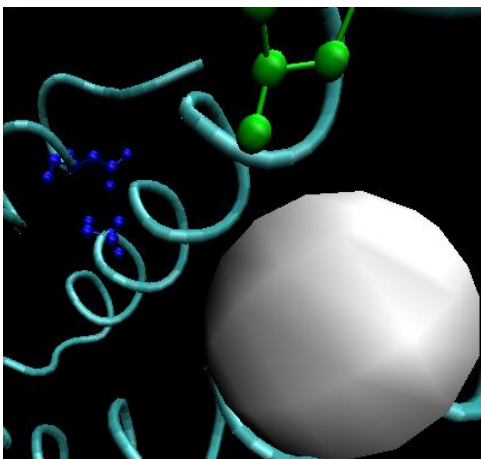
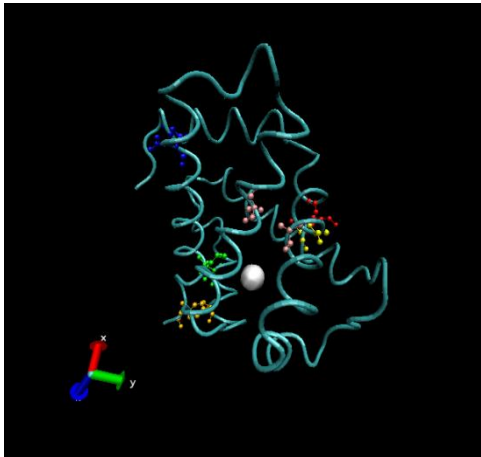
D'aquí se'n pot treure una conclusió interessant: tots els aminoàcids d'una mateixa hèlix  $\alpha$  tindran una interacció forta, encara que la distància que els separi sigui considerable.

### ***Zones de forta interacció fora de la línia central***

Com ja hem dit, que els aminoàcids que estan molt a prop en la seqüència tinguin una interacció forta no és sorprenent. Ara bé, també trobem zones d'interacció forta fora de la línia central, les quals provoquen un seguit de taques més vermelloses per la quadrícula. L'estudi d'aquestes zones és, per tant, interessant. Així que el que vam fer va ser seleccionar el parell d'aminoàcids amb la interacció més forta de cada zona i situar-lo a l'estructura terciària amb el VMD, també agafant l'hemoglobina humana d'exemple.

Els parells en qüestió resulten ser els següents: Asp 6 - Ser 124 (blau), Gly 22 - Lys 60 (vermell), Ala 26 - Gly 59 (groc), Ala 28 - Leu 105 (rosa), Ser 84 - Leu 136 (verd), i Ala 88 - Tyr 140 (taronja).

A les dues primeres imatges podem veure dues perspectives de l'hemoglobina humana amb totes els parells d'aminoàcids ressaltats. Per tal de veure de més a prop cada parell, llavors tenim tot d'imatges on es poden veure els parells de més a prop: primer l'Asp 6 i la Ser 124, després la Gly 22 i la Lys 60, i l'Ala 26 i la Gly 59, seguits per l'Ala 28 i la Leu 105 i, finalment, el la Ser 84 i la Leu 136, i l'Ala 88 i la Tyr 140.



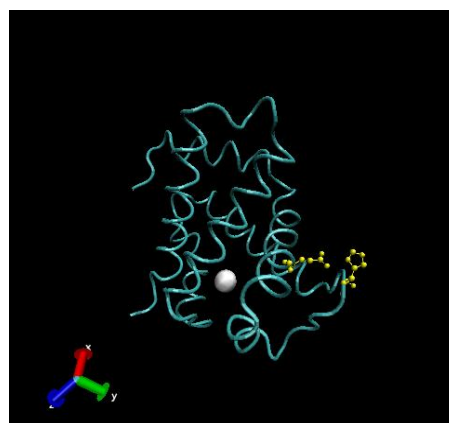
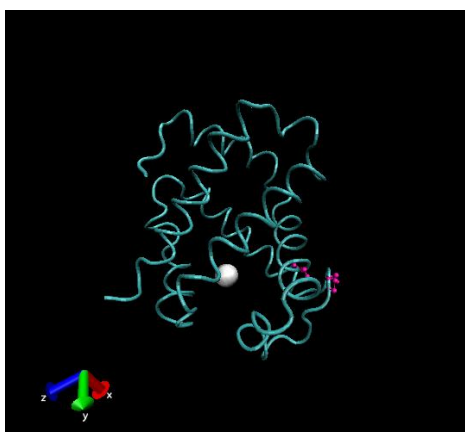
Com ens ha passat en el cas dels quadrats, aquí la explicació a la interacció forta també és clara: tot i que els dos aminoàcids que formen un parell es trobin molt separats en l'estructura primària, estan situats molt a prop en l'estructura terciària, gairebé tocant-se. D'aquí el seu alt grau d'interacció.



### Comparació de l'Apparent Stiffness entre les diferents espècies

Si comparem la quadrícula de l'Apparent Stiffness entre les diferents espècies, no podem destacar diferències molt evidents. Podem dir que l'hemoglobina del gos és la que té, en general, tons més foscos, però poca cosa més perquè les espècies tenen les mateixes zones d'interacció forta fora de la línia central. Ara bé, si ens hi fixem més en detall, podem veure que la tonyina té un parell d'aminoàcids amb interacció forta que no té cap altra espècie: el parell Gly 31 - Pro 52, el qual correspondria a Glu 30 - His 50 en les altres espècies.

Si ressaltem el parell Gly 31 - Pro 52 en la tonyina (1<sup>a</sup> imatge) i el parell Glu 30 - His 50 en l'humà (2<sup>a</sup> imatge) utilitzant el VMD, la resposta també és clara, com es pot veure en les imatges de sota: mentre que l'hemoglobina humana té la part inferior dreta de la cadena de l'estructura terciària molt separada de la cadena principal, la tonyina la té gairebé enganxada. Això fa que la tonyina tingui un altre parell d'aminoàcids amb interacció forta, ja que aquests queden molt junts. En l'humà, en canvi, el parell queda a una distància considerable.



### Aminoàcids amb interacció alta amb el grup hemo

Un altre aspecte interessant de l'Apparent Stiffness és saber amb quins aminoàcids el grup hemo hi té més interacció. Per escollir quins eren aquests aminoàcids, el valor que havien de superar era el 2%. Aquests són els aminoàcids obtinguts:

<b>Humà</b>	Val 62, His 87 i Gly 88
<b>Gos</b>	Gly 22, Asp 23, Lys 60, Val 62, Ala 63, His 87, Ala 88 i Tyr 89
<b>Ànec de collverd</b>	Glu 22, Glu 23, Val 62, His 87, Ala 88 i Gln 89
<b>Colom</b>	Val 62, Ala 63, His 87, Ala 88 i Gln 89
<b>Tonyina</b>	Asp 23, Lys 62, Val 64, Met 65, His 89, Ala 90 i Phe 91

Els aminoàcids del quadre es poden classificar en tres grups, separats per la distància en l'estructura primària. En cada grup hi consta un color, el qual serà amb el que estaran pintades quan situem els diferents grups en l'estructura terciària:

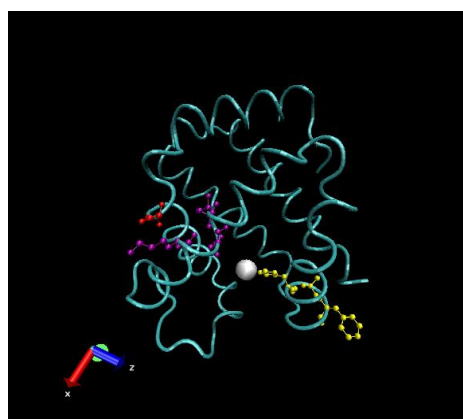
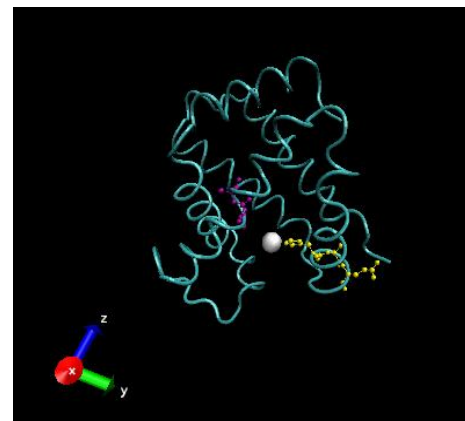
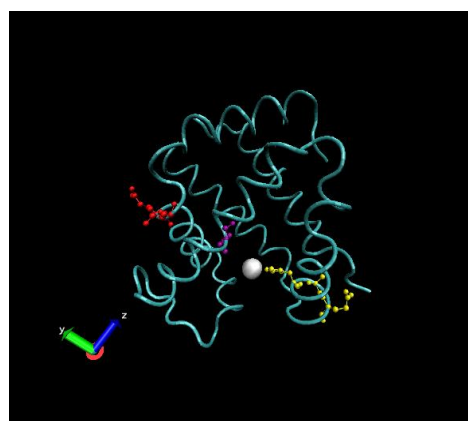
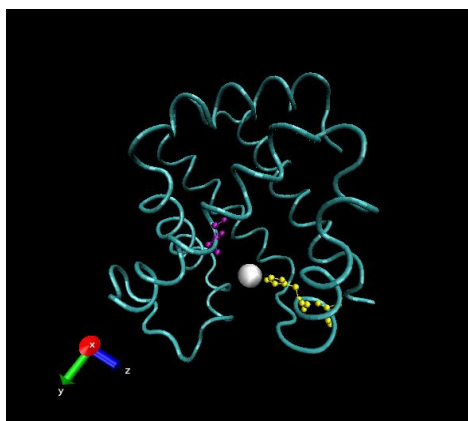
- Primera (vermella): 22 i 23
- Segona (lila): 60, 62, 63, 64 i 65
- Tercera (groga): 87, 89, 90, 91



Com es pot veure, totes les espècies coincideixen bastant en els aminoàcids, si tenim en compte que per saber l'equivalent dels aminoàcids de la tonyina per les altres espècies, hem de restar un aminoàcid al primer grup de la tonyina, i dos en els altres grups, ja que una deleció o inserció es va produir entre els grups primer i segon. Mirant a la taula podem pensar que els grups hemo de l'hemoglobina humana i de la del colom no tenen interacció amb el primer grup. Sí que en tenen, i força destacable si la comparem amb les interaccions tan baixes que expressen amb la resta d'aminoàcids, però no passava del 2%.

També veiem que el gos és l'espècie amb més aminoàcids destacats. De fet, ja havíem vist a la quadrícula de l'*Apparent Stiffness* que el gos era l'espècie amb el valor més alts en general.

Les imatges següents corresponen, per aquest ordre i d'esquerre a dreta, a l'hemoglobina de l'humà, del gos, de l'ànec de collverd, del colom i de la tonyina.



Un cop observades, és senzill entendre la interacció forta amb els segon i tercer grups, ja que aquests estan molt a prop del grup hemo. Fins i tot un aminoàcid del tercer grup, l'His 87 (His 89 en la tonyina), està gairebé tocant el grup hemo representat per la glicina, i d'aquí que tongui els valors d'interacció més alts, superior al 3% en totes les espècies menys en l'humà i el colom.

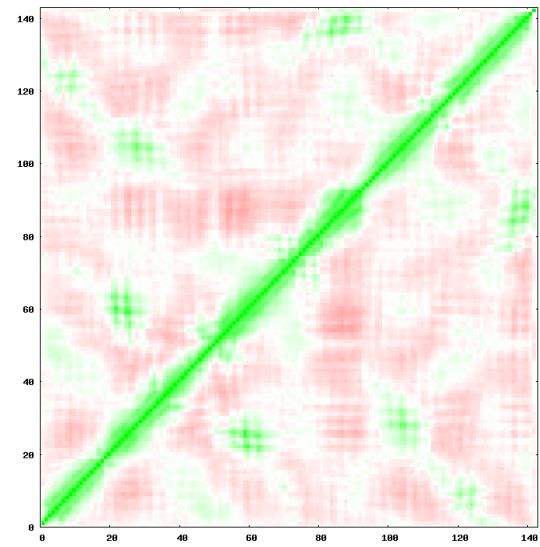
Ara bé, trobar l'explicació de la forta interacció amb el primer grup pot semblar més complicada, ja que la posició d'aquest està molt allunyada del grup hemo. De fet, és més senzill del que sembla si sabem com funciona l'*Apparent Stiffness*. Com que el grup hemo té una interacció forta amb el segon grup i aquest últim amb el primer perquè estan molt a prop, l'*Apparent Stiffness* conclou que el grup hemo també té interacció destacable amb el primer grup. No calcula les interaccions només directament, sinó que també té en compte aquest tipus d'interaccions.

Com que aquests aminoàcids tenen una interacció alta amb el grup hemo, podríem pensar que s'han conservat al llarg del temps i que, per tant, són comuns en totes les espècies. Podem comprovar-ho mirant la taula, ja que cada número d'aminoàcid és acompanyat pel nom de l'aminoàcid. Per mirar-ho amb més detall, però, podem consultar l'alineament realitzat amb el ClustalW2.

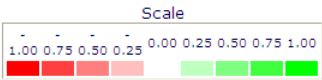
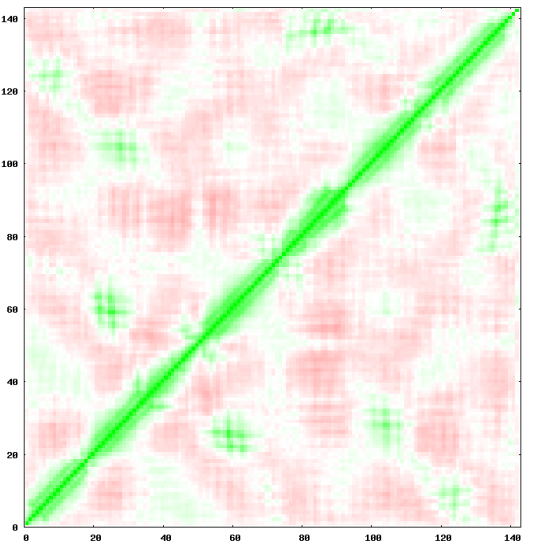
El primer grup no presenta conservació, ja que, com es pot veure, els aminoàcids varien entre les espècies. Tampoc és tan estrany perquè la interacció amb el grup hemo és indirecta. Pels dos altres grups, en canvi, no estaríem tan allunyats de la nostra hipòtesi, ja que el segon es troba en part en una de les zones perfectament conservades que hem estudiat anteriorment, i el tercer, tot i no estar en una d'aquestes zones, el trobem situat també en part sobre una zona també molt conservada. Així, pot ser que aquesta forta interacció hagi fet que no es fixessin mutacions en els dos últims grups.

*Residue correlation*

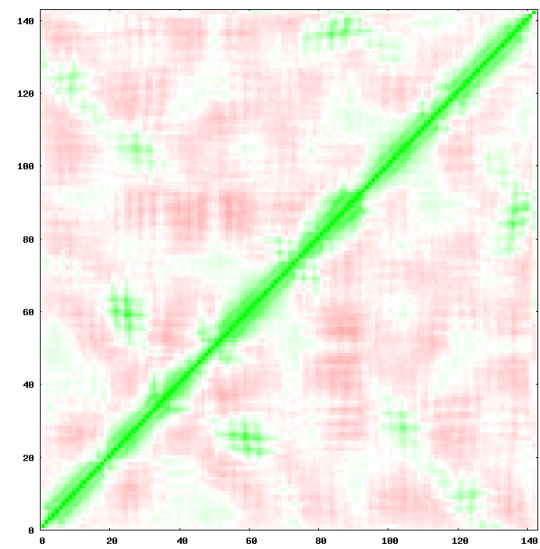
Humà



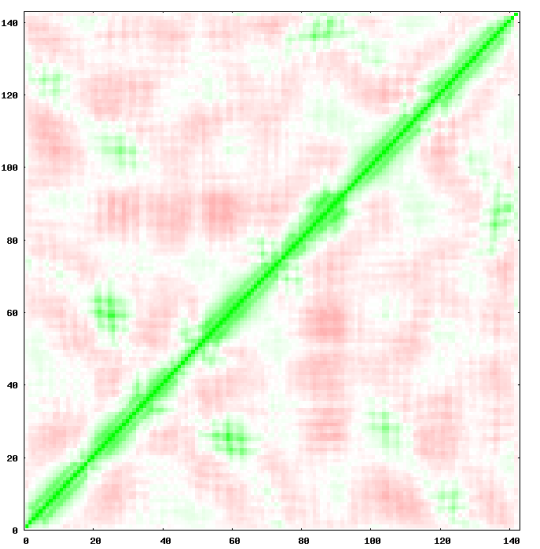
Gos



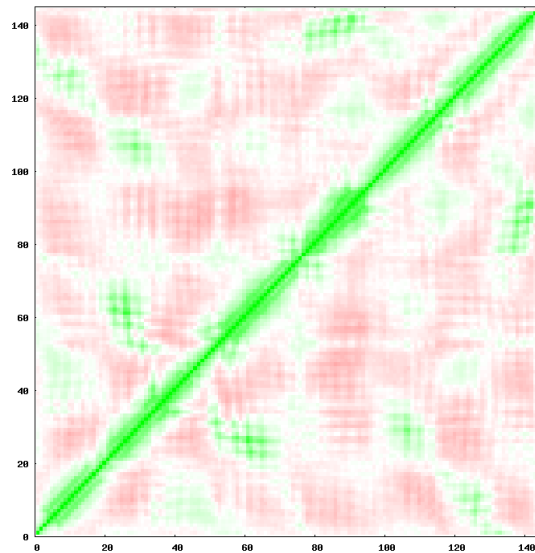
Ànec de collverd



Colom



## Tonyina

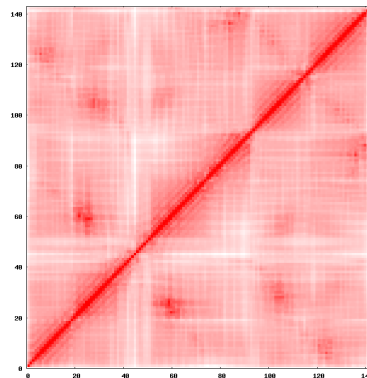
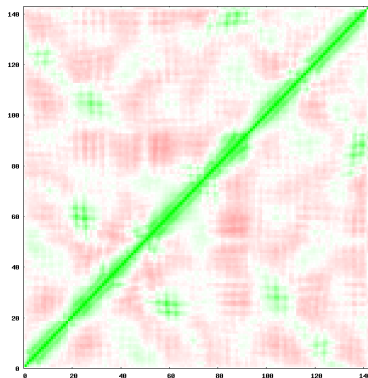


### Comentaris

Pel mateix motiu que trobàvem en l'*Apparent Stiffness*, la quadrícula del *Residue correlation* també és simètrica, on la línia verda que la travessa fa d'eix. És normal que els aminoàcids propers en l'estructura primària tinguin una correlació forta, ja que si un es mou, els altres es canvien necessàriament de posició. També podem veure que la línia verda es torna més prima en certs punts, els quals corresponen als mateixos que trobàvem en l'*Apparent Stiffness*. O sigui, que també indiquen el principi o final d'una hèlix  $\alpha$ .

### **Zones verdes**

A part de la línia verda, també trobem taques verdes escampades per la quadrícula. Les altres zones tenen un color lleugerament vermellós. Comparant la quadrícula humana del *Residue correlation* amb la corresponent de l'*Apparent Stiffness* es veu clarament que les zones verdes de la primera coincideixen amb les de vermell fort de la segona, com es pot comprovar amb les imatges següents:



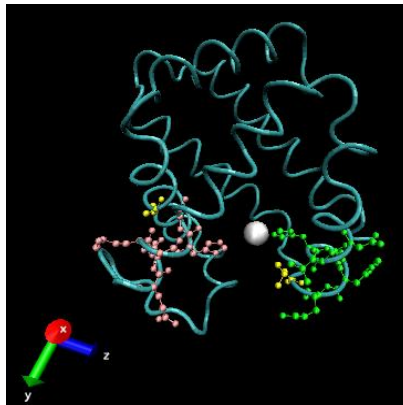
A més, després de buscar el parell d'aminoàcids de cada zona verda amb una correlació més forta, vam trobar exactament els mateixos aminoàcids que havíem destacat en l'apartat de l'*Apparent Stiffness*. D'aquesta manera, sabem que els parells d'aminoàcids que formen les zones verdes són parells d'aminoàcids que es troben allunyats en l'estructura primària, però a prop en la terciària. Per tant, aquests aminoàcids no tan sols tenen una interacció forta, sinó que el moviment d'un afectaria la posició de l'altre.

### ***Zones vermelloses***

Pel que fa a les zones vermelloses, les zones amb més correlació difereixen una mica segons l'espècie. D'aquí que cada espècie tingui un apartat propi. Per considerar una correlació forta, vaig agafar els valors de -0'3% a -1% (el màxim).

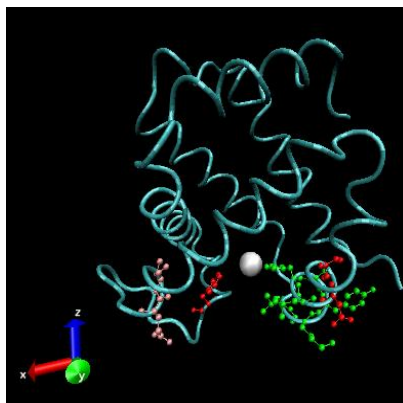
#### *Humà*

En l'hemoglobina humana trobem valors alts en el parell Ala 26 - Lys 91 (groc) i, seguidament, grans correlacions entre la zona de la Gln 54 a Gly 59 (rosa) i la zona de l'Asp 85 a la Leu 91 (marró).



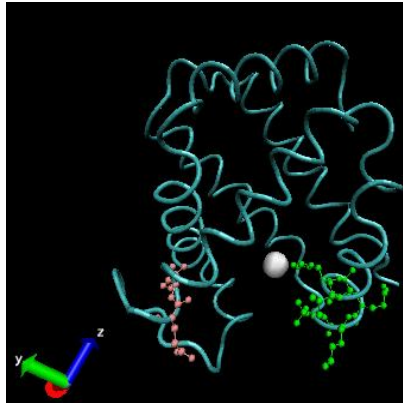
#### *Gos*

En l'hemoglobina del gos trobem menys correlació. Comencem amb una de forta de la Phe 43 amb la Ser 84 i l'Asp 85 (vermell) i, per últim, els aminoàcids de la Leu 86 a la Leu 91 (verd) amb l'Asp 54 i la Val 55 (rosa).



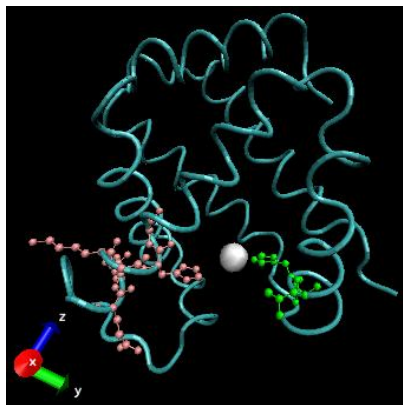
### *Ànec de collverd*

L'hemoglobina de l'ànec de collverd, les correlacions més importants són les de la zona compresa entre la Leu 86 i la Leu 91 (verd) amb els aminoàcids Asp 54 i Ile 55 (rosa).



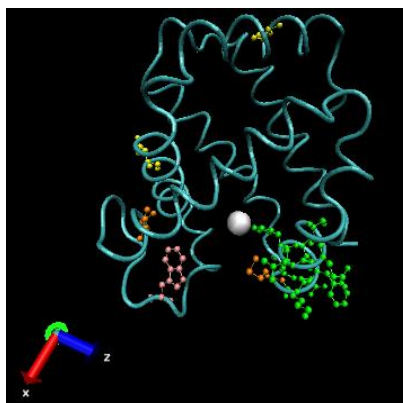
### *Colom*

L'hemoglobina del colom presenta una correlació destacable entre la zona de la Gln 54 a Gly 59 (rosa) i els aminoàcids Leu 86 i His 87 (verd).



### *Tonyina*

L'hemoglobina de tonyina té altes correlacions en els parells d'aminoàcids Thr 10 -Asp 28 (groc) i el Val 57 - Met 93 (taronja). A més, els aminoàcids de la Ser 86 a l'Arg 94 (verd) presenten una correlació alta amb els aminoàcids Trp 47 (rosa) i Val 57.



## Comentaris

A diferència de les zones verdes, on els parells d'aminoàcids estaven molt a prop en l'estructura terciària, els aminoàcids de les zones vermelloses no hi estan. De totes maneres, això no torna erronis els nostres resultats de correlació, ja que no necessàriament dos aminoàcids han d'estar a prop per estar correlacionats.

Les correlacions potser més fàcils d'explicar són les existents entre els aminoàcids 50 i següents i els aminoàcids 80 i següents, les quals són comunes en totes les espècies. Si mirem la situació d'aquestes dues zones en l'estructura terciària veurem que es troben en hèlix  $\alpha$  adjacents. Per tant, fàcilment si una hèlix es mou, farà moure l'altra. De fet, aquestes zones resulten tenir totes dues un gran moviment en el B-factor.

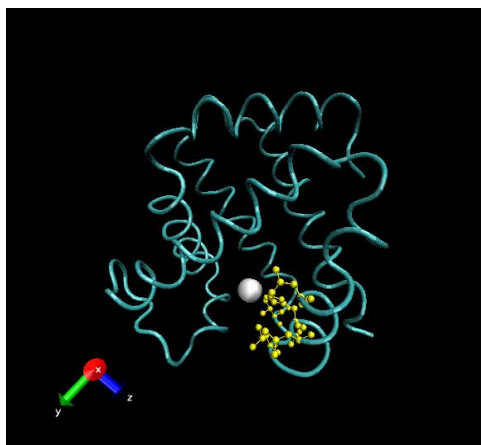
### **Aminoàcids amb correlació alta amb el grup hemo**

Per últim, ens falta estudiar els aminoàcids amb alta correlació amb el grup hemo, de manera que, si aquests es mouen, el grup hemo també ho farà. El valor mínim considerat és de 0'15%.

<b>Humà</b>	Leu 83, Leu 86, His 87, Leu 91 i Val 93
<b>Gos</b>	Leu 83, Leu 86, His 87 i Leu 91
<b>Ànec de collverd</b>	Leu 83, Leu 86, His 87 i Leu 91
<b>Colom</b>	Leu 83, Leu 86, His 87 i Val 93
<b>Tonyina</b>	Leu 85, Leu 88, His 89, Met 93, Val 95

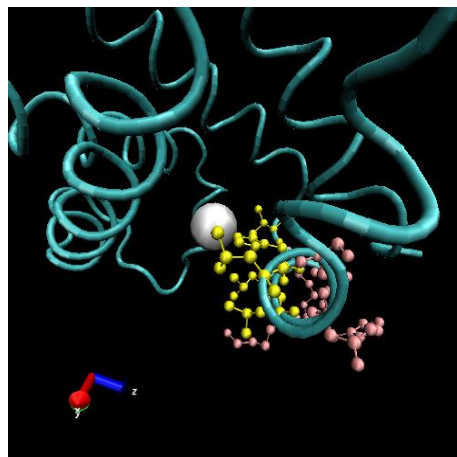
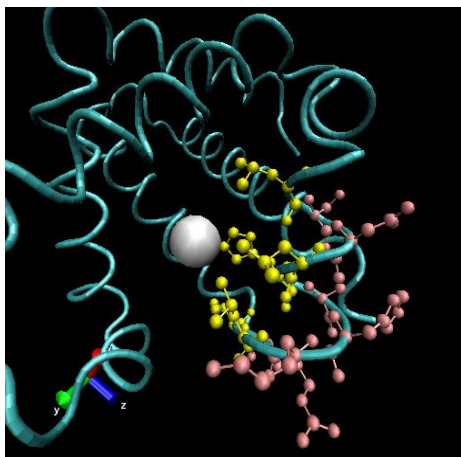
Sorprenentment, totes les espècies coincideixen exactament amb els mateixos aminoàcids, recordant sempre que la tonyina ha de tenir dos aminoàcids més que la resta de les espècies. L'única diferència és que l'humà, el colom i la tonyina tenen un aminoàcid més que les altres, la Val 93/95, tot i que al colom li falta la Leu 91, la qual arribava a 0'14% i no a 0'15%. Fins i tot el tipus de l'aminoàcid és el mateix, amb l'excepció de la Met 93, que correspon a la Leu 91. Així veiem que la conservació d'aquests aminoàcids és molt i molt alta.

Com que els aminoàcids són tan semblants per a totes les espècies, agafem l'hemoglobina humana com a exemple per a representar-los.



Aquesta imatge ens situa els aminoàcids (de color groc) a l'estructura terciària. Efectivament, estan molt a prop del grup hemo. Tot i així, aquesta imatge no ens permet entendre com és que en totes les espècies tenen exactament els mateixos aminoàcids, els quals van saltats, i

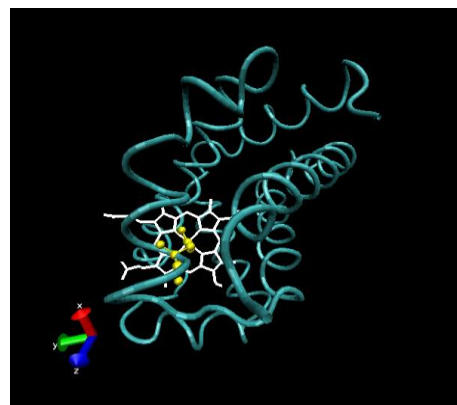
no els que es torben entre aquests. Per tal de poder-ho descobrir, a les següents imatges hi trobem els aminoàcids de la taula anterior de color groc igualment, però també hi trobem representats de rosa els aminoàcids que es troben intercalats entre els anteriors. Els aminoàcids en qüestió són la Ser 84, l'Asp 85, l'Ala 88, l'His 89, la Lys 90 i l'Arg 92. Les imatges resultants són les següents:



Els resultats són contundents: els aminoàcids amb la correlació més forta amb el grup hemo són els aminoàcids que, tot i estar dins la mateixa hèlix  $\alpha$ , estan encarats al grup hemo, sense cap excepció. Efectivament, és normal que estiguin molt conservats, ja que fins i tot, aquests aminoàcids tenen els seus residus gairebé tocant la glicina que representa el grup hemo.

### Histidina pròxima

L'aminoàcid amb la correlació més alta amb el grup hemo és l'His 87/89, la qual arriba a un valor de 0'22% o superior en totes les espècies. Aquesta Histidina també havíem trobat que era l'aminoàcid amb la interacció més forta amb el grup hemo. Aquests valors són d'esperar, ja que el residu d'aquest aminoàcid gairebé toca la Glicina que representa el grup hemo. Tot i així, aquesta Glicina és simplement un  $C\alpha$  amb unes mides superiors a les reals per tal de ressaltar-lo. Així, per fer-nos una idea de la situació d'aquest aminoàcid en l'hemoglobina real, hem de tornar a utilitzar el fitxer pdb sense la modificació del grup hemo.



Utilitzant altre cop l'hemoglobina humana com a exemple, veiem a la primera imatge la distància entre l'His 87, de groc, i el grup hemo. En la segona imatge veiem que, curiosament, el residu de la Histidina està situat perpendicularment respecte el grup hemo i, sobretot, perfectament situat sobre el ferro del grup hemo. Aquest fet pot significar que aquesta Histidina tingui alguna funció relacionada amb el funcionament del grup hemo.

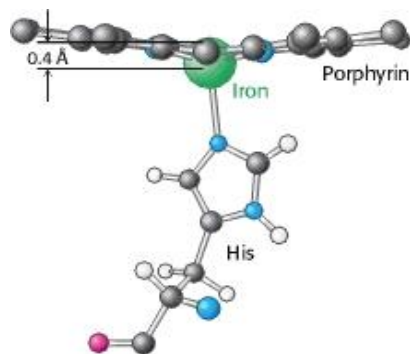


### Funció de la Histidina pròxima

Després de consultar diferent bibliografia, no queda cap dubte de la importància de la Histidina 87/89 i que segurament no se seleccionarà mai positivament una mutació en aquest aminoàcid. El ferro pot formar sis enllaços: quatre els dedica als quatre nitrògens que el subjecten, un a la Histidina 87/89 i l'altra a l'oxigen que se l'hi ha d'unir. Aquesta Histidina, anomenada "Histidina pròxima", connecta el ferro a la cadena polipeptídica, de manera que el ferro ara és sensible als canvis de l'estructura de l'hemoglobina i viceversa.

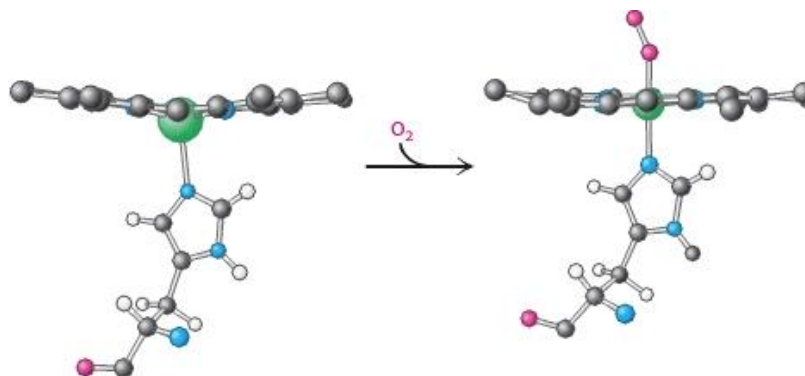
D'aquesta manera, aquesta connexió provoca canvis en l'estructura terciària de l'hemoglobina, els quals poden transmetre els seus efectes a les altres cadenes polipeptídiques. Així, que l'oxigen s'uneixi a una cadena afecta indirectament l'afinitat de les altres cadenes a l'oxigen. L'hemoglobina funciona de la següent manera: cada oxigen que s'uneix, incrementa la possibilitat que a les altres cadenes també se n'uneixin.

El ferro, quan no està unit a l'oxigen, es troba situat a una distància de 0'4 Å del grup hemo perquè el ferro és lleugerament massa gran per encaixar a l'espai tan ben definit pels quatre nitrògens (Imatge 10).



Imatge 10: Grup hemo de costat amb el ferro de color verd unit a la Histidina pròxima (Font: NCBI)

La unió de la molècula d'oxigen al sisè enllaç del ferro provoca que els electrons del ferro es reagrupin, de manera que el ferro esdevé més petit i, per tant, ja pot situar-se al mateix pla que la resta del grup hemo (Imatge 11). A la Imatge 11 també podem veure que els àtoms de la Histidina pròxima també canvien de posició un cop l'oxigen s'ha unit al ferro. Aquest canvi és el que es transmetrà a les altres cadenes, fent que aquestes tinguin més facilitat per unir-se a una altra molècula d'oxigen.



Imatge 11: Grup hemo unit a la Histidina pròxima i grup hemo unit a la Histidina pròxima i a una molècula d'oxigen (Font: NCBI)

## Discussió

L'objectiu d'aquest apartat era analitzar com l'evolució havia actuat en hemoglobines de diferents espècies animals. Penso que el fet que l'hemoglobina fos la proteïna estudiada ha estat una bona elecció, ja que té una estructura terciària molt senzilla i també una zona catalítica clarament diferenciada, el grup hemo, cosa que en facilitava l'estudi.

Al principi de l'apartat només hi consta informació molt general i bàsica de l'hemoglobina, ja que l'objectiu no era estudiar el funcionament d'aquesta proteïna, sinó tan sols estudiar-ne l'evolució. Tot i així, amb els resultats que hem obtingut en el subapartat del FlexServ, es veia clar que s'havia d'investigar més a fons la Histidina 87/89, de manera que al final hi trobem un altre apartat teòric dedicat a aquest aminoàcid i que ens permet conèixer molt millor el funcionament del grup hemo i el transport d'oxigen.

Pel que fa a la comparació de les cinc proteïnes estudiades, hagués estat bé tenir un mínim d'un representant per grup de vertebrats. Per desgràcia, al banc de dades del PDB no tenien cap hemoglobina de rèptil o d'amfibi. I de peixos només vaig poder-hi trobar una espècie coneguda, la de la tonyina. De fet, hi havia més peixos, i pot ser que també hi hagués algun rèptil i algun amfibi, però preferia treballar només amb espècies molt conegudes. Tot i així, les cinc espècies escollides van anar molt bé a l'hora de realitzar l'arbre filogenètic d'aquestes. A més, els resultats esperats van acabar sent els obtinguts.

On sí que em vaig equivocar va ser en els resultats esperats de l'estructura primària i terciària de les hemoglobines. Jo m'havia esperat trobar diferències importants entre elles en l'estructura terciària, però molt poques en l'estructura primària, ja que em pensava que alguns canvis en la seqüència d'aminoàcids ja provocarien forces alteracions en l'estructura terciària. Els resultats obtinguts van acabar sent el contrari: molts aminoàcids diferents i, en canvi, aparences estructurals gairebé idèntiques.

L'estudi de les zones perfectament conservades i de la zona amb molt poca conservació ha sigut útil per veure que la selecció natural té molt en compte si una mutació es produeix en una zona catalítica de la proteïna, o en un lloc proper, en el nostre cas. Hem vist que les zones perfectament conservades es trobaven a prop dels vèrtexs del grup hemo, mentre que la zona amb molt poca conservació la situàvem en la zona més allunyada del grup hemo. L'estudi de la situació de les mutacions de la zona poc conservada en l'arbre filogenètic l'he realitzat només per especular sobre el moment evolutiu de les mutacions i sobretot veure si aquestes eren possibles amb una sola mutació puntual.

Pel que fa a l'estudi del FlexServ ens ha servit per conèixer més a fons les hemoglobines amb les que hem estat treballant, tant el seu moviment, com les seves interaccions i correlacions. I, comparant els resultats dels diferents factors, també hem pogut veure les diferències entre elles. S'ha de dir que les diferències entre elles eren molt petites, cosa que és normal perquè el que té en compte el FlexServ és l'estructura terciària que, com ja hem dit, era molt similar en totes les espècies. Aquests resultats també ens han servit per preguntar-nos sobre la funció de la Histidina pròxima, la qual ha resultat tenir una funció molt important per a l'hemoglobina.

El que ens quedaria per preguntar-nos i que es podria realitzar en un altre estudi és comparar com es comporten les diferents hemoglobines en presència de molècules d'oxigen, ja que potser els canals per on passa l'oxigen són diferents segons l'espècie. Aquest estudi potser ens permetria veure més diferències entre les diferents espècies.

# Conclusions

---

Les mutacions es poden definir tan sols dient “canvis a l’atzar en el genoma”, i poden ser provocades tan per canvis en l’ADN freqüents en genoma com per agents externs. La gran varietat en els factors que les provoquen comporta un ventall ampli de diferents tipus de mutacions, on cada una té efectes diferents, sent perjudicials, silencioses o beneficioses per a l’organisme que les pateix. El fet de recordar que són a l’atzar és important perquè, com s’accepta actualment en el món científic, les mutacions no tenen una direcció, o sigui, no es produeixen per complir una finalitat específica dins l’organisme.

Si parlem en general, podem afirmar que les mutacions són d’una gran importància per a assegurar la permanència d’una espècie a la Terra. Com que canvien el genoma dels organismes, provoquen una gran diversitat dins d’una mateixa espècie. Així, si els factors ambientals canvien, és probable que hi hagi un cert nombre d’individus d’una espècie que, gràcies a una o diverses mutacions, puguin suportar les noves condicions. Aquests darrers individus seran seleccionats positivament per la selecció natural, mentre que la resta ho seran negativament. I d’aquesta manera és com les espècies evolucionen.

Les mutacions, a part de propiciar l’evolució de les espècies, també provoquen que se’n creïn de noves, donant lloc a l’especiació. Les espècies estan constituïdes per moltes poblacions d’aquesta espècie. Com que les mutacions són a l’atzar, és lògic que cada població presenti mutacions diferents. Si dues poblacions arriben al punt de no poder-se encreuar ni tenir descendència fèrtil, es poden considerar que són dues espècies diferents. Podem representar tots els processos d’especiació que s’han produït a partir d’un avantpassat comú mitjançant els arbres filogenètics, que ens ajuden en l’estudi de la diversitat biològica i l’evolució.

Tan les mutacions com l’evolució són fets propis de la naturalesa. Ara bé, també es poden reproduir artificialment per tal d’estudiar l’evolució? En el segon apartat d’aquest treball científic hem vist que sí. Hem comprovat que mutacions a l’atzar en un gen poden produir que n’obtinguem un de nou que codifiqui una nova proteïna funcional.

Així, modificant la proteïna HisF mitjançant la error-prone-PCR per tal d’introduir-hi mutacions, i vam obtenir un gen que vam anomenar hisF\*, que havia de ser capaç de codificar per una proteïna que pogués dur a terme la funció de la HisA. Aconseguir-ho no era difícil perquè la HisF i la HisA són molt semblants. Mitjançant diferents procediments, vam obtenir un plasmidi amb el gen hisF\*. El vam introduir en bacteris sense el gen hisA, i aquests en una placa de Petri amb un medi mínim per tal de sotmetre’ls a la selecció natural.

Una placa amb els bacteris que tenien el gen hisF\* va presentar creixement, cosa que significava que havien suplert la manca de la proteïna HisA amb una altra proteïna que podia fer la seva funció, i aquesta era la HisF\*. No en podia ser cap altre perquè els bacteris que no tenien el plasmidi amb el gen hisF\* no van sobreviure. Aquest experiment ens demostra que les mutacions a l’atzar poden donar lloc a gens nous que codifiquin proteïnes funcionals. I és mitjançant aquest sistema com s’obtenen noves funcions dins d’una població, i d’aquí s’esdevé la diversitat.

Un cop vist com evolucionen les proteïnes, només quedava estudiar-ne un cas real. L'estudi de les hemoglobines va resultar ser molt interessant. Hem vist que la mateixa proteïna és diferent segons l'espècie a la qual pertany. Aquestes diferències les hem vist principalment en les seqüència d'aminoàcids, ja que els canvis en l'estructura terciària eren molt petits. Així, resulta que moltes diferències en l'estructura primària no sempre significa grans canvis en l'estructura terciària de les proteïnes.

Va ser precisament amb l'estudi de les seves seqüències com vam veure quines eren més semblants amb quines, i d'aquí vam establir-ne relacions evolutives entre les diferents espècies. Ha resultat que l'hemoglobina humana i la del gos estan molt a prop evolutivament, igual que la de l'ànec de collverd i el colom. Aquests dos grups, que es poden generalitzar dient mamífers i ocells, tenen un avantpassat comú i aquest últim amb la tonyina. Així, mamífers i ocells estan més a prop evolutivament entre ells que no pas amb la tonyina, que representaria els peixos.

Pel que fa a l'estudi de les zones perfectament conservades, o sigui, els grups d'aminoàcids comuns en totes les espècies, hem vist que la seva posició a l'estructura terciària tenia una raó de ser; totes se situaven prop dels vèrtexs del grup hemo, cosa que és lògic perquè canvis en aquesta zona poden dificultar la funció de la part catalítica de la molècula. Així, es veu clar la importància d'aquestes zones per a l'hemoglobina, ja que qualsevol animal que ha patit una mutació en aquest punt no ha sobreviscut, ja que cap mutació no s'ha fixat a aquestes parts de l'hemoglobina i, per tant, no ha passat a les espècies actuals.

Per últim, vam estudiar la flexibilitat de les diferents proteïnes mitjançant el FlexServ, i hem pogut conèixer el seu moviment, les seves interaccions i les seves correlacions. Les diferències entre elles eren molt poc notòries perquè el FlexServ té en compte l'estructura terciària, que era molt semblant entre elles.

Com a conclusió general, es pot dir que hem vist palès el paper que tenen les mutacions en l'evolució i l'especiació, la manera com es produeixen les mutacions i són seleccionades positivament o negativament, i que l'estudi d'una mateixa proteïna present en diferents espècies ens proporciona molta informació per entendre els mecanismes de l'evolució.

# Agraïments

---

En primer lloc, m'agradaria destacar i agrair l'ajuda que m'ha ofert l'Annabel Villarejo, professora de Biologia a l'Institut Montsacopa i tutora del meu treball de recerca, que m'ha ajudat a encaminar el meu treball, m'ha donat consell i s'ha hagut de llegir i rellegir el meu treball.

Per la part teòrica de *Com les mutacions porten l'evolució i l'especiació*, m'agradaria citar la tasca que van realitzar en Sergi Aranda i l'Ariadna Laguna, que van ser els meus professors del projecte de Genètica i Biologia Molecular a les Planes de Son, i que van ser els encarregats d'aconsellar-me i supervisar l'article científic que vaig utilitzar per aquest apartat.

Per l'apartat de *l'Evolució artificial de la proteïna HisF*, vull agrair a la Martina Henn-Sax, professora del projecte *Molecular Biology* que vaig cursar durant la primera setmana de la meva estada a l'XLAB, haver-me proporcionat les imatges dels resultats obtinguts i certa informació extra.

Arribats a aquest punt, no puc deixar de citar la beca *Estades d'Estiu de Ciència de Caixa Catalunya* del programa Joves i Ciència de Caixa Catalunya, ja que gràcies a la beca he pogut realitzar els dos projectes que m'han permès la realització dels dos apartats anteriors, a part d'haver-me permès viure experiències inoblidables relacionades amb el món científic i haver-me obert les portes d'aquest món.

Per últim, per l'apartat *Evolució i comparació de l'hemoglobina en diferents espècies* vaig comptar amb el magnífic ajut del Parc Científic de Barcelona gràcies a la beca *Recerca a Secundària* de Caixa Catalunya, i en especial de l'Oliver Carrillo, el meu tutor en el PCB. Tenia una gran llibertat a l'hora d'enfocar i dirigir el meu treball, però sempre comptant amb el suport de l'Oliver, que em va ajudar sobretot en el subapartat relacionat amb el FlexServ. Ell em va explicar en quina teoria es basaven tots els programes que vaig utilitzar i com usar-los. També vam poder discutir sobre els resultats obtinguts en aquest subapartat i em va parlar de la importància de la Histidina pròxima.

I no em puc deixar de citar el suport dels meus pares que, tot i no acabar d'entendre de què anava el meu treball, sempre m'han sabut donar bonsconsell.

Gràcies a tots.

# Bibliografia

---

## Bibliografia de *Com les mutacions porten l'evolució i l'especiació*

- Awise, J. C. i F. J. Ayala (2009). *In the light of evolution III: Two centuries of Darwin* (Department of Ecology and Evolutionary Biology, University of California, Irvine).
- Baum, D. (2008). *Reading a phylogenetic tree: The meaning of monophyletic groups*. Nature Education 1(1). (Dept. of Botany, University of Wisconsin). Disponible des d'Internet a: <<http://www.nature.com/scitable/topicpage/reading-a-phylogenetic-tree-the-meaning-of-41956>> [Consultat al 2009].
- Campbell, N. A. i J. B. Reece (2007). *Biología*. Madrid, Espanya. Ed: Panamericana.
- Clancy, S. (2008a). *Genetic mutation*. Nature Education 1(1). Disponible des d'Internet a: <<http://www.nature.com/scitable/topicpage/genetic-mutation-441>> [Consultat al 2009].
- Clancy, S. (2008b). *DNA damage & repair: mechanisms for maintaining DNA integrity*. Nature Education 1(1). Disponible des d'Internet a: <<http://www.nature.com/scitable/topicpage/dna-damage-repair-mechanisms-for-maintaining-dna-344>> [Consultat al 2009].
- Duret, L. (2008). *Neutral theory: The null hypothesis of molecular evolution*. Nature Education 1(1). (Laboratoire de Biométrie et Biologie Évolutive, Université Claude Bernard, France). Disponible des d'Internet a: <<http://www.nature.com/scitable/topicpage/neutral-theory-the-null-hypothesis-of-molecular-839>> [Consultat al 2009].
- Jimeno, A. i M. Ballesteros (2009). *Biología 1*. Barcelona, Espanya. Ed: Grup Promotor Santillana.
- Jimeno, A. i L. Ugedo (2009). *Biología 2*. Barcelona, Espanya. Ed: Grup Promotor Santillana.
- Kutschera, U. (2009). *Charles Darwin's Origin of Species, directional selection, and the evolutionary sciences today* (Institute of Biology, University of Kassel).
- Loewe, L. (2008). *Genetic mutation*. Nature Education 1(1) (School of Biological Sciences, University of Edinburgh, Scotland, UK). Disponible des d'Internet a: <<http://www.nature.com/scitable/topicpage/genetic-mutation-1127>> [Consultat al 2009].
- Mayr, E. (1963). *Animal Species and Evolution* (The Belknap, Cambridge, Massachusetts).
- McClintock, B. (1950). *The origin and behavior of mutable loci in maize*. Proceedings of the National Academy of Sciences.
- McClintock, B. (1953). *Induction of instability at selected loci in maize*. Genetics.
- Pray, L. (2008a) *Transposons: The jumping genes*. Nature Education 1(1). Disponible des d'Internet a: <<http://www.nature.com/scitable/topicpage/transposons-the-jumping-genes-518>> [Consultat al 2009].

Pray, L. (2008b). *DNA replication and causes of mutation*. Nature Education 1(1). Disponible des d'Internet a: <<http://www.nature.com/scitable/topicpage/dna-replication-and-causes-of-mutation-409>> [Consultat al 2009].

Saitou, N. i M. Nei (1987). *The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees*. (Center for Demographic and Population Genetics, the University of Texas. Health Science Center at Houston).

Scitable, de Nature education: <<http://www.nature.com/scitable>> [Consultat al 2009].

Strachan, T. i A. P. Read (1999). *Genética molecular humana*. Barcelona, Espanya. Ed: Omega.

TalkOrigins Archive. Exploring the Creation/Evolution Controversy: <<http://www.talkorigins.org>> [Consultat al 2009].

Wu, C. and Ting, C. (2004). *Genes and Speciation* (C. Wu: Department of Ecology and Evolution, University of Chicago, Chicago i C. Ting: Department of Life Science, National Tsing Hua University, Hsinchu).

## **Bibliografia d'Evolució artificial de la proteïna HisF**

Henn-Sax, M. (2008). *Evolution in the test tube*. XLAB- Experimentallabor für junge Leute e.V. Document lliurat en el projecte de Molecular Biology realitzat a l'XLAB del 21 al 25 de juny. S'hi explica tot el necessari per a la realització de l'experiment de la HisA i la HisF. Per tant, tota la informació teòrica i pràctica d'aquest experiment té la font en aquest document i en la informació afegida per la professora del projecte, la Dra. Martina Henn-Sax.

Lang, D., R. Thoma, M. Henn-Sax, R. Sterner i M. Wilmanns (1 de setembre del 2000). *Structural Evidence for Evolution of the  $\beta/\alpha$  Barrel Scaffold by Gene Duplication and Fusion*.

## **Bibliografia d'Evolució i comparació de l'hemoglobina en diferents espècies**

ClustalW2: Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H. \*, Valentin F. \*, Wallace I.M., Wilm A., Lopez R. \*, Thompson J.D., Gibson T.J. and Higgins D.G. (2007) ClustalW and ClustalX version 2. *Bioinformatics* 23(21): 2947-2948. Disponible des d'Internet a: <<http://www.ebi.ac.uk/Tools/clustalw2>> [Consultat al 2010].

FlexServ: <<http://mmb.pcb.ub.es/FlexServ>> [Consultat al 2010]. Aquí s'hi pot trobar tant el programa com tota la informació teòrica de la introducció i l'explicació dels diferents factors estudiats.

Jalview: Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. and Barton, G. J. (2009) "Jalview Version 2 - a multiple sequence alignment editor and analysis workbench" *Bioinformatics* 25 (9) 1189-1191 doi: 10.1093/bioinformatics/btp033.

National Center for Biotechnology Information (NCBI): <<http://www.ncbi.nlm.nih.gov>> [Consultat al 2010].

Protein Data Bank (PDB): <<http://www.rcsb.org/pdb/home/home.do>> [Consultat al 2010].

Proteopedia, life in 3D: <<http://www.proteopedia.org>> [Consultat al 2010].