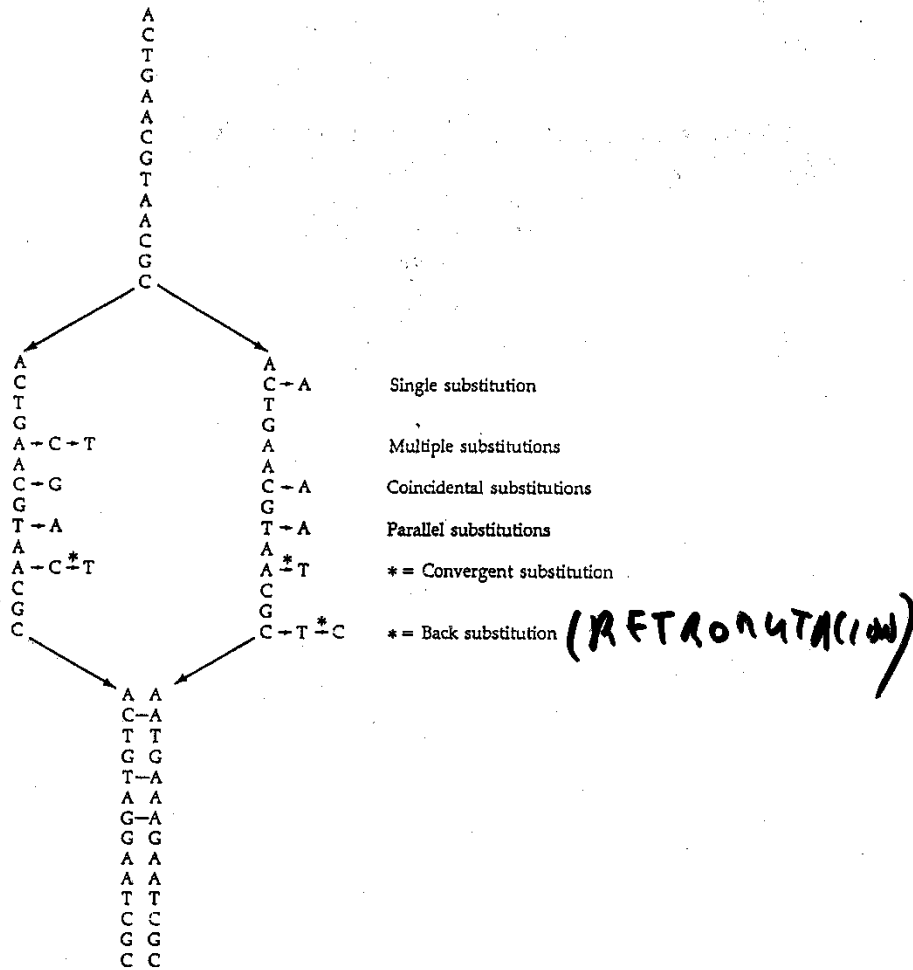


Evolución de secuencias de proteínas

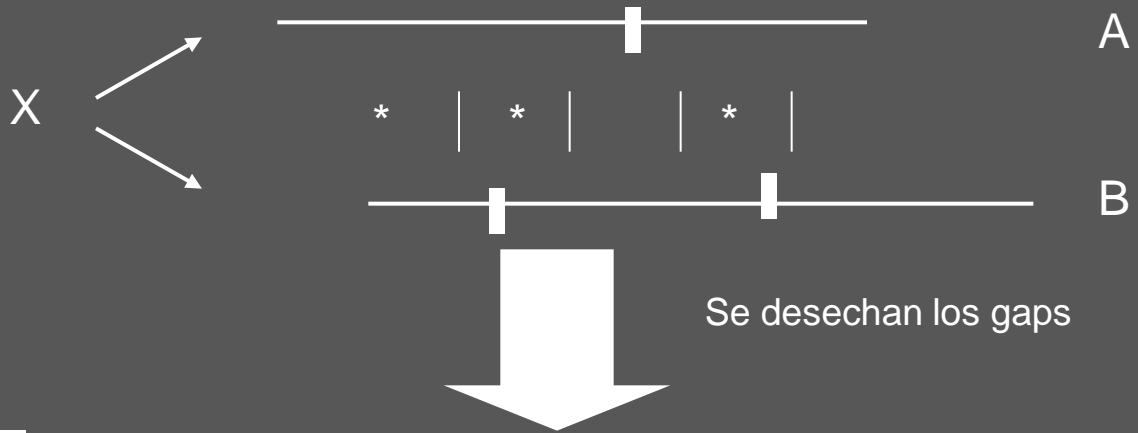
Sustituciones múltiples



12 mutaciones

→ Solo 3 cambios

Figure 5. Two homologous DNA sequences which descended from an ancestral sequence and accumulated mutations since their divergence from each other. Note that although 12 mutations have accumulated, differences can be detected at only three nucleotide sites. Note further that "coincidental substitutions," "parallel substitutions," "convergent substitutions," and "back substitutions" all involve multiple substitutions at the same site, though maybe in different lineages.



n_d

Número de posiciones con Aa diferentes

n_i

“ “ “ “ “ idénticos

$$n_d + n_i = n$$

$$p = n_d / n$$

con lo que

$$p + q = 1$$

$$q = n_i / n$$

p es la distancia o divergencia evolutiva cruda (sin corregir)



A: Arg **Pro** Lys Leu --- Phe Pro...

B: Arg **Glu** Lys Leu His Phe Pro ...

$$p = n_d / n$$

Si p es pequeño \rightarrow es aproximadamente igual al número de sustituciones de Aa por sitio

Si p es grande \rightarrow es una estima por defecto del número de sustituciones de Aa por sitio

\rightarrow Se necesita un modelo matemático: distribución de Poisson

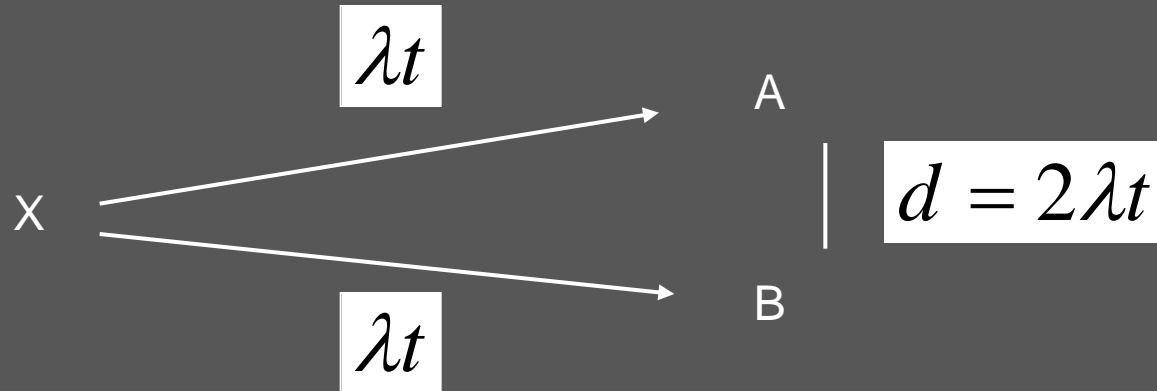
[muchos sucesos (20), todos ellos con probabilidad pequeña (1/19)]

λ

Tasa de sustitución de Aa por año y por sitio

 t

Tiempo en años



La probabilidad de que en la secuencia ancestral X ocurran r sustituciones en un sitio en t años viene dada por la distribución de Poisson:

$$P_r(t) = e^{-\lambda t} \cdot (\lambda t)^r / r!$$

La probabilidad de que durante t años no haya ocurrido ninguna sustitución, es decir, si r = 0

$$P_0(t) = e^{-\lambda t} \cdot (\lambda t)^0 / 0!$$

es decir

$$P_0(t) = e^{-\lambda t}$$

Así pues, la probabilidad (q) de que no haya ocurrido ninguna sustitución en ninguna de las dos secuencias (sucesos independientes) será:

$$q = e^{-\lambda t} \bullet e^{-\lambda t} = e^{-2\lambda t}$$

tomando logaritmos

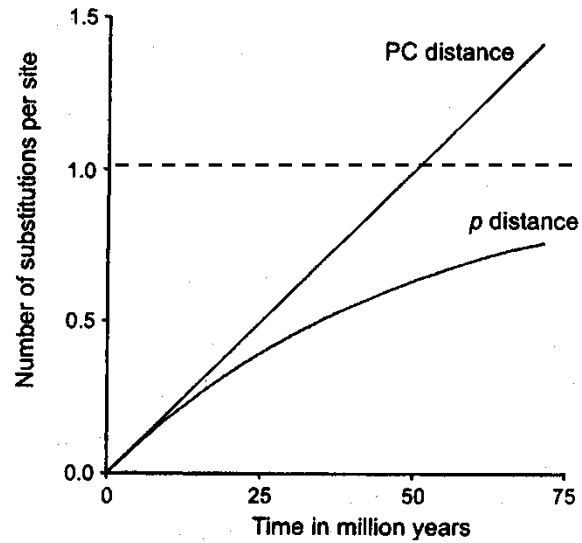
$$\ln q = -2\lambda t$$

y puesto que $d = 2\lambda t$

$$d = -\ln q$$

Distancia evolutiva

PC (Corrección de Poisson)



$$2\lambda t = 1$$

FIGURE 2.2. Relationships of the p distance and the Poisson correction (PC) distance with time. The rate of amino acid substitution (λ) is assumed to be 10^{-8} per site per year.

1	0	1	1	0	0	1	1	0	0
1	1	0	1	0	0	1	1	0	1

3 diferencias



1	0	1	0	1	0	0	1	1	0
0	1	0	1	0	1	1	0	0	1

10 diferencias



Saturación

EXAMPLE

Figure 4.2 shows the amino acid sequences of hemoglobin α chains from the human, horse, bovine, and carp. Here, amino acids are represented by one-letter codes rather than by usual three-letter codes (see table 3.2). The three mammalian hemoglobins consist of 141 amino acids, whereas the carp hemoglobin has 142 amino acids. Comparison of these sequences suggests that deletions or insertions occurred at three different positions after the divergence between fish and mammals. If we ignore these deletions/insertions, the proportion of different amino acids

Human	VLSPADKTNVKAANGKVG AHAGEYGA EALERMFLSFPTTKTYFPHF-DLSHGSAQVKGHG
Horse	VLSAADKTNVKAAWSKVGGHAGEYGA EALERMFLGFPTTKTYFPHF-DLSHGSAQVKAHG
Bovine	VLSAADKGNVKAANGKVGHAAEYGA EALERMFLSFPTTKTYFPHF-DLSHGSAQVKGHG
Carp	SLSDKDKA AVKIAWAKISPKADDIGAEALGRMLTVYPQTKTYFAHWADLSPGSGPVK-HG
Human	KKVA-DALTNVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP AEFT
Horse	KKVA-DGLTLAVGHLDDLPGALSDLSNLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFT
Bovine	AKVA-AALTKAVEHLDDLPGALSELSDLHAHKLRVDPVNFKLLSHSLLVTLASHLPDFT
Carp	KKVINGAVGDAVSKIDDLVGGLASLSELHASKLRVDPANFKILANHIVVGIMFYLP GDFP
Human	PAVHASLDKFLASVSTVLT SKYR
Horse	PAVHASLDKFLSSVSTVLT SKYR
Bovine	PAVHASLDKFLANVSTVLT SKYR
Carp	PEVHMSVDKFFQNLALAL SEKYR

Figure 4.2. Amino acid sequences in the α chains of hemoglobins from four vertebrate species. Amino acids are expressed in terms of one-letter codes. The hyphens indicate the positions of deletions or insertions.

	Hombre	Caballo	Vaca	Carpa
Hombre	--	18	17	68
Caballo	0.138	--	18	66
Vaca	0.129	0.138	--	65
Carpa	0.666	0.637	0.624	--

Hombre-Carpa:

$$p = 68 / 140 = 0.486$$



Distancia cruda

$$d = -\ln q = -\ln (1-0.486) = 0.666$$



Distancia corregida (PC)

	Hombre	Caballo	Vaca	Carpa
Hombre	--	18	17	68
Caballo	0.138	--	18	66
Vaca	0.129	0.138	--	65
Carpa	0.666	0.637	0.624	--

$$0.642 / 0.135 \approx 5$$

$$400 \text{ MA} / 75 \text{ MA} \approx 5$$

$$D = -\ln q$$

$$V_d = (1-q) / (qn)$$

Secuencias cortas → Varianza muy grande

$$\begin{matrix} x_{11}, & x_{12}, & x_{13}, & x_{14}, & x_{15}, & \dots, & x_{1n} \\ x_{21}, & x_{22}, & x_{23}, & x_{24}, & x_{25}, & \dots, & x_{2n} \end{matrix} \quad (S2.1)$$



Muestreo de sitios
(posiciones, columnas) al
azar con remplazamiento
(bootstrapping)

$$\begin{matrix} x_{11}, & x_{12}, & x_{12}, & x_{14}, & x_{15}, & \dots, & x_{1n-1} \\ x_{21}, & x_{22}, & x_{22}, & x_{24}, & x_{25}, & \dots, & x_{2n-1} \end{matrix} \quad (S2.2)$$

Algunas posiciones
están repetidas

... y otras faltan, como x_{13}

El proceso se repite 100 ó
1000 veces y se calcula la
varianza

EVOLUTIONARY CHANGE OF AMINO ACID SEQUENCES

Table 2.3 Standard error estimates of the PC distances by the analytical (below diagonal) and bootstrap (above diagonal) methods.

	Human	Horse	Cow	Kangaroo	Newt	Carp
Human		0.031	0.031	0.039	0.078	0.083
Horse	0.031		0.030	0.043	0.083	0.081
Cow	0.031	0.031		0.038	0.080	0.079
Kangaroo	0.040	0.043	0.039		0.081	0.084
Newt	0.074	0.080	0.076	0.080		0.090
Carp	0.082	0.081	0.079	0.086	0.089	

Hemos supuesto que la tasa de sustitución de Aa es la misma en todas las posiciones.

Esto no siempre es cierto, ya que dicha tasa suele ser más alta en los sitios menos importantes desde el punto de vista funcional.

Las variaciones de la tasa de sustitución de Aa en diferentes posiciones sigue una distribución gamma, lo que permite corregir la distancia PC.

Se ha observado experimentalmente que existe un mayor número de sustituciones entre Aa bioquímicamente similares que entre Aa de distintos grupos bioquímicos.

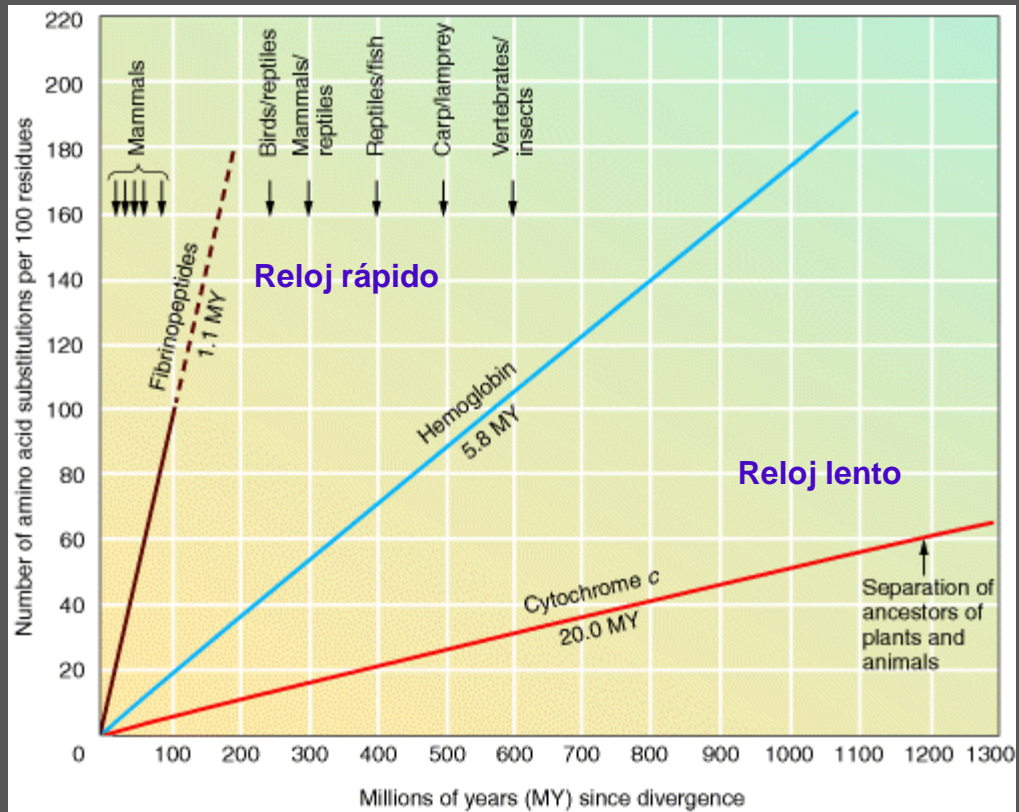
Algunos Aa (cisteína, glicina, triptófano) cambian muy raramente, mientras que otros lo hacen más a menudo.

Esto también puede alterar la estima de la distancia.

Dayhoff propuso por tanto otro método para calcular la distancia, basado en las llamadas matrices PAM (Point Accepted Mutations).

		Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	A	9867	2	9	10	3	8	17	21	2	6	4	2	6	2	22	35	32	0	2	18
Arg	R	1	9913	1	0	1	10	0	0	10	3	1	19	4	1	4	6	1	8	0	1
Asn	N	4	1	9822	36	0	4	6	6	21	3	1	13	0	1	2	20	9	1	4	1
Asp	D	6	0	42	9859	0	6	53	6	4	1	0	3	0	0	1	5	3	0	0	1
Cys	C	1	1	0	0	9973	0	0	0	1	1	0	0	0	0	1	5	1	0	3	2
Gln	Q	3	9	4	5	0	9876	27	1	23	1	3	6	4	0	6	2	2	0	0	1
Glu	E	10	0	7	56	0	35	9865	4	2	3	1	4	1	0	3	4	2	0	1	2
Gly	G	21	1	12	11	1	3	7	9935	1	0	1	2	1	1	3	21	3	0	0	5
His	H	1	8	18	3	1	20	1	0	9912	0	1	1	0	2	3	1	1	1	4	1
Ile	I	2	2	3	1	2	1	2	0	0	9872	9	2	12	7	0	1	7	0	1	33
Leu	L	3	1	3	0	0	6	1	1	4	22	9947	2	45	13	3	1	3	4	2	15
Lys	K	2	37	25	6	0	12	7	2	2	4	1	9926	20	0	3	8	11	0	1	1
Met	M	1	1	0	0	0	2	0	0	0	5	8	4	9874	1	0	1	2	0	0	4
Phe	F	1	1	1	0	0	0	0	1	2	8	6	0	4	9946	0	2	1	3	28	0
Pro	P	13	5	2	1	1	8	3	2	5	1	2	2	1	1	9926	12	4	0	0	2
Ser	S	28	11	34	7	11	4	6	16	2	2	1	7	4	3	17	9840	38	5	2	2
Thr	T	22	2	13	4	1	3	2	2	1	11	2	8	6	1	5	32	9871	0	2	9
Trp	W	0	2	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	9976	1	0
Tyr	Y	1	0	3	0	3	0	1	0	4	1	1	0	0	21	0	1	1	2	9945	1
Val	V	13	2	1	1	3	2	2	3	3	57	11	1	17	1	3	2	10	0	2	9901

Table 1: Margaret Dayhoff's 1978 mutation data matrix for 1 PAM. Frequencies have been multiplied by 10000. For a given column, the rows give the probabilities of change, so that each column adds up to 1.



Excepciones:

- Insulina
- Super Oxido dismutasa (SOD)

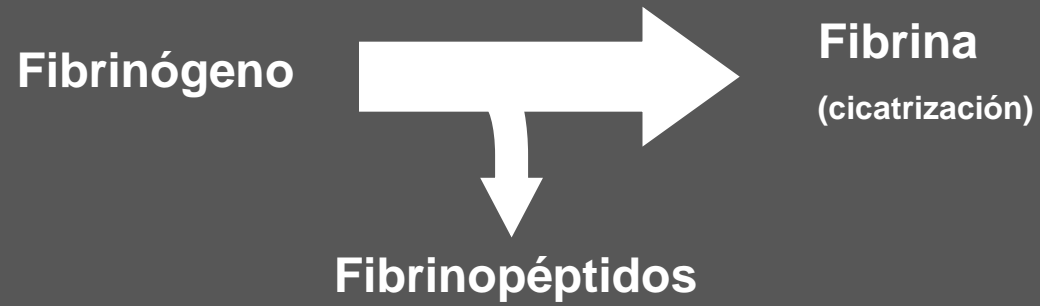
Dickerson: Number of amino acid substitutions in the evolution of the vertebrates as a function of time since divergence. The three proteins—fibrinopeptides, hemoglobin, and cytochrome c—differ in rate because different proportions of their amino acid substitutions are selectively neutral.

Table 4.4 Rates of amino acid substitutions per amino acid site per 10^9 years ($\lambda \times 10^9$) in various proteins. Modified from Dayhoff (1978).

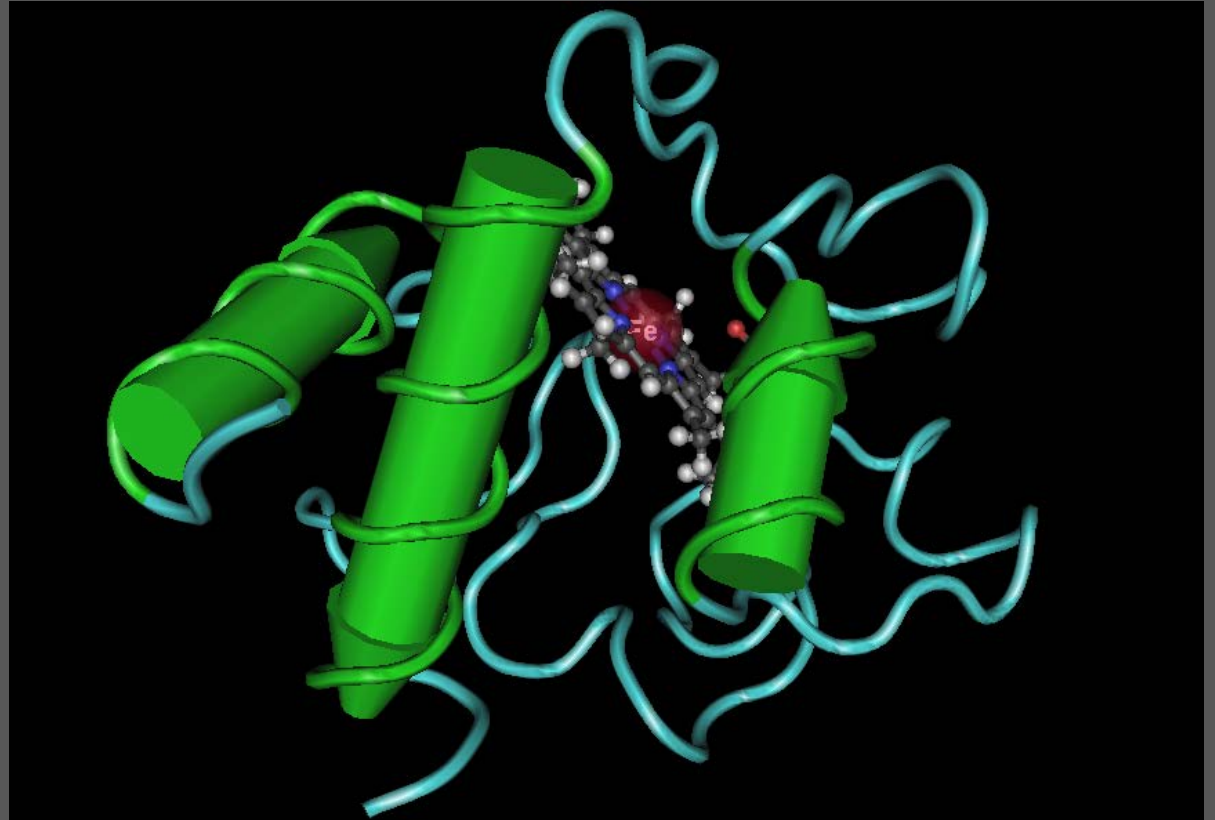
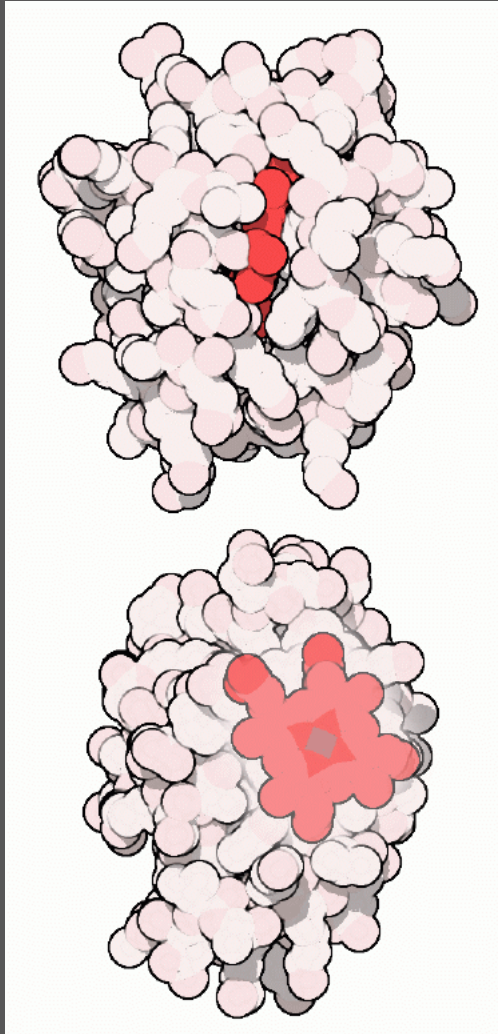
Protein	Rate	Protein	Rate
Fibrinopeptides	9.0	Thyrotropin beta chain	0.74
Growth hormone	3.7	Parathyrin	0.73
Ig kappa chain C region	3.7	Parvalbumin	0.70
Kappa casein	3.3	Protease inhibitors, BP1 type	0.62
Ig gamma chain C region	3.1	Trypsin	0.59
Lutropin beta chain	3.0	Melanotropin beta	0.56
Ig lambda chain C region	2.7	Alpha crystallin A chain	0.50
Complement C3a anaphylatoxin	2.7	Endorphin	0.48
Lactalbumin	2.7	Cytochrome b ₅	0.45
Epidermal growth factor	2.6	Insulin (exc. guinea pig and coypu)	0.44
Somatotropin	2.5	Calcitonin	0.43
Pancreatic ribonuclease	2.1	Neurophysin 2	0.36
Lipotropin beta	2.1	Plastocyanin	0.35
Haptoglobin alpha chain	2.0	Lactate dehydrogenase	0.34
Serum albumin	1.9	Adenylate kinase	0.32
Phospholipase A ₂	1.9	Triosephosphate isomerase	0.28
Protease inhibitor, PST1 type	1.8	Vasoactive intestinal peptide	0.26
Prolactin	1.7	Corticotropin	0.25
Pancreatic hormone	1.7	Glyceraldehyde 3-PO ₄ dehydrogenase	0.22
Carbonic anhydrase C	1.6	Cytochrome c	0.22
Lutropin alpha chain	1.6	Plant ferredoxin	0.19
Hemoglobin alpha chain	1.2	Collagen (exc. nonrepetitive ends)	0.17
Hemoglobin beta chain	1.2	Troponin C, skeletal muscle	0.15
Lipid-binding protein A-II	1.0	Alpha crystallin B chain	0.15
Gastrin	0.98	Glucagon	0.12
Animal lysozyme	0.98	Glutamate dehydrogenase	0.09
Myoglobin	0.89	Histone H2B	0.09
Amyloid AA	0.87	Histone H2A	0.05
Nerve growth factor	0.85	Histone H3	0.014
Acid proteases	0.84	Ubiquitin	0.010
Myelin basic protein	0.74	Histone H4	0.010

Thus, T_e for fibrinopeptides is 1.1×10^6 years, whereas T_e for histone H4 is 10^9 years.

La unidad de tiempo evolutivo (T_e) es el tiempo medio requerido para que haya 1 sustitución por cada 100 posiciones



Citocromo C



Solution Structure Of Reduced Recombinant Human Cytochrome C



Bar-Headed Goose
Hemoglobin (Oxy Form)

Crystal Structure Of Human
Hemoglobin A2

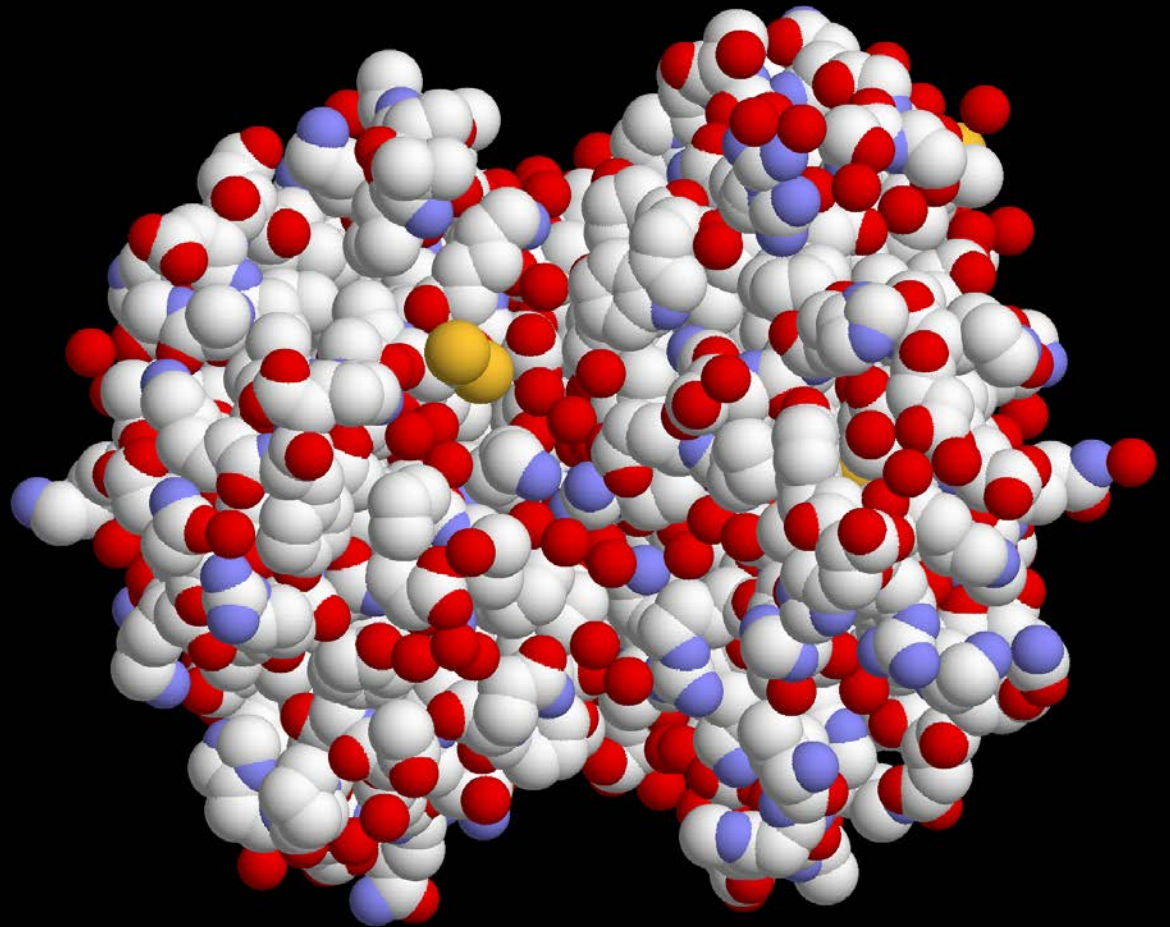




Figure 8.9. Two views of the nucleosome core octamer. The view on the left is downwards from the top of the barrel-shaped octamer; the view on the right is from the side. The two strands of the DNA double helix wrapped around the octamer are shown in brown and green. The octamer comprises a central tetramer of two histone H3 (blue) and two histone H4 (bright green) subunits plus a pair of H2A (yellow) H2B (red) dimers, one above and one below the central tetramer. Note the N-terminal tails of the histone proteins protruding from the core octamer. Reprinted with permission from Luger *et al.*, *Nature*, 389, 251-260. Copyright 1997 Macmillan Magazines Limited.

En los fibrinopéptidos

$$\lambda = u$$

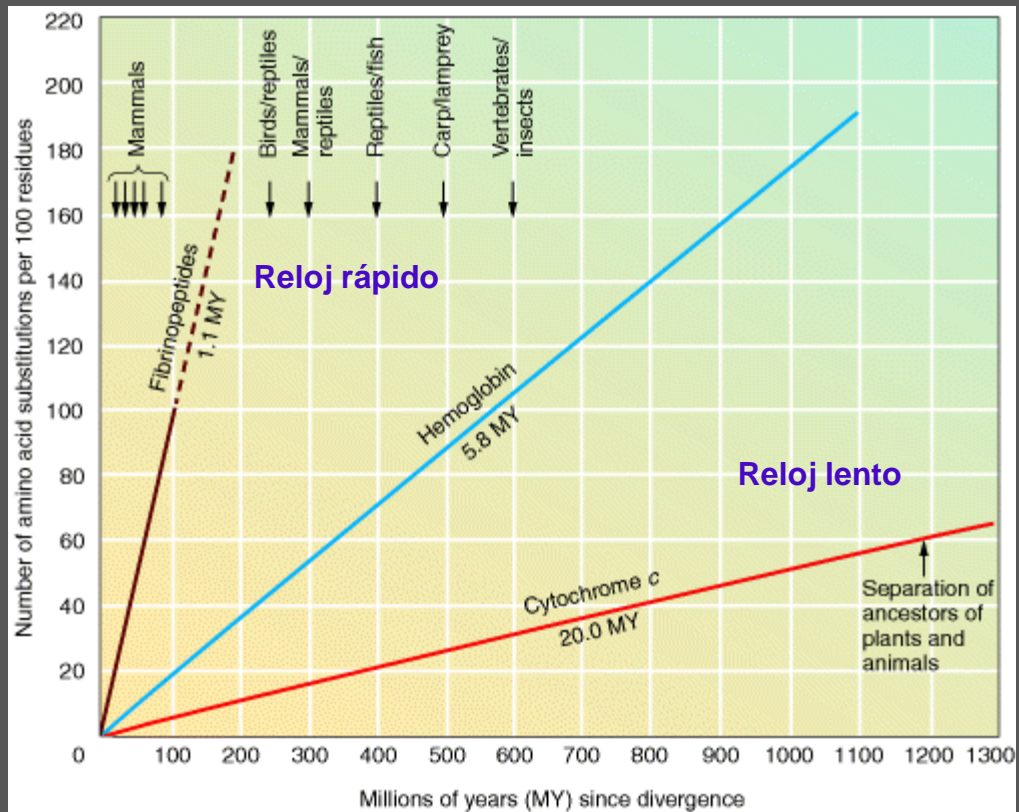
λ : Tasa de sustitución por sitio

u : Tasa de mutación

En el resto de las proteínas

$$\lambda = (1-f) u$$

f : Proporción de
mutaciones deletéreas



Dickerson: Number of amino acid substitutions in the evolution of the vertebrates as a function of time since divergence. The three proteins — fibrinopeptides, hemoglobin, and cytochrome c— differ in rate because different proportions of their amino acid substitutions are selectively neutral.

¿Todo se debe entonces a selección negativa?

¿Es que no existen mutaciones ventajosas?

La respuesta es que sí, pero las mutaciones ventajosas son mucho menos frecuentes



Ejemplo: Hb de los cocodrilos

- Ha perdido su función de unirse al CO_2
- Sin embargo, se une al grupo bicarbonato



Respuesta adaptativa a la mayor acidez de la sangre de los cocodrilos

(pasan largos periodos bajo el agua -> hipoxia -> acidez)

HBB_SQUAC/7-137	LTVA ^Y NHI.....D.DV ^K PHFV ^L SKK.HYEEL...HVDPENFKLLANCL
HBB_HETPO/7-136	LGVA ^V THL.....G.DVKSQFTDLSKK.HAEEL...HVDVESFKLLAKCF
HBB_RANCA/1-135	IEE ^G LK ^H F.....Z.BLKAYYAKLSER.HS ^G EL...HVDPANFYRLGNVL
HBB1_XENBO/6-140	IGEAIKHM.....D.DIKGYAQLSKY.HSETL...HVDPCNFKRFGGCL
HBB_LEPPA/7-142	IIECT ^R HF.....G.NIKGHLANLSHL.HSEKL...HVDPHNFRVLGQCI
HBB1_CYGMA/7-141	LDR ^G LKNM.....D.NIVDAYAELSTL.HSEKL...HVDPDNFKLLSDCI
HBB_ALLMI/7-141	FGEAVKHL.....D.NIKGHFANLSKL.HCEKF...HVDPENFKLLGDII
HBBN_AMMLE/2-136	FSB ^G LKHL.....B.BLKGAFASLSZL.HCBKL...HVBPZBFRLLG ^B VL
HBB0_MOUSE/7-141	LGLAVKNM.....D.NLKETF ^A HLSL.HCDKL...HADPENFKLLGNML
HBB2_XENLA/7-141	VGNAISHI.....D.SVKSSLQQLSKI.HATEL...FVDPENFKRFGGVL
HBA_SQUAC/6-136	VADAT ^D H ^LD.NVAGHLDPLAVL.HG ^T T ^L ...CVDPHNFP ^L LLTQCI
HBA1_TORMA/6-136	LVD ^T NKHL.....D.DLPHHLNKLAEK.HG ^K GL...LVDPHNFKLFSDCI
HBA_HETPO/13-143	LAKACDHL.....D.DLKTHLHKLATF.HS ^G EL...KVDPANFQYLSYCL
HBA_LEPPA/7-138	IGQGVQHL.....H.DLSSCLHTLSEK.HAREL...MVDPCNFQY ^L IEAI
HBA_LIOMI/6-136	LDSALKHL.....D.DIR ^G TLSQLSDL.HAYNL...RVDPVNFQLLSKCI
HBA1_PASMO/6-136	LSNAIKNL.....D.NLSQALSEL ^S NL.HAYNL...RVDPVNFKFLSQCL

Alineamiento de la globina de los cocodrilos con la humana:

- El cambio adaptativo se debe a sustituciones de Aa en tan solo 5 posiciones
- Pero hay un total de 123 sustituciones entre las globinas de ambas especies
- La mayoría de las mutaciones no están relacionadas con esta adaptación

El cambio evolutivo está dominado por las mutaciones neutras

→ Reloj molecular

Modificación de la Hb humana para conseguir el mismo incremento en el transporte de oxígeno:

- Terapia de enfermos con enfisema, etc.
- Mejora del rendimiento de los atletas?

Conclusión: basta un pequeño número de sustituciones para provocar un cambio adaptativo.

Otros ejemplos:

- **Cambios en la especificidad de sustrato o resistencia a antibióticos en bacterias.**
- **Los mutantes resistentes a penicilina muestran una menor afinidad por los antibióticos, lo que se debe tan solo a 4-5 sustituciones de Aa.**