



Elementos de físico-química en biomedicina I

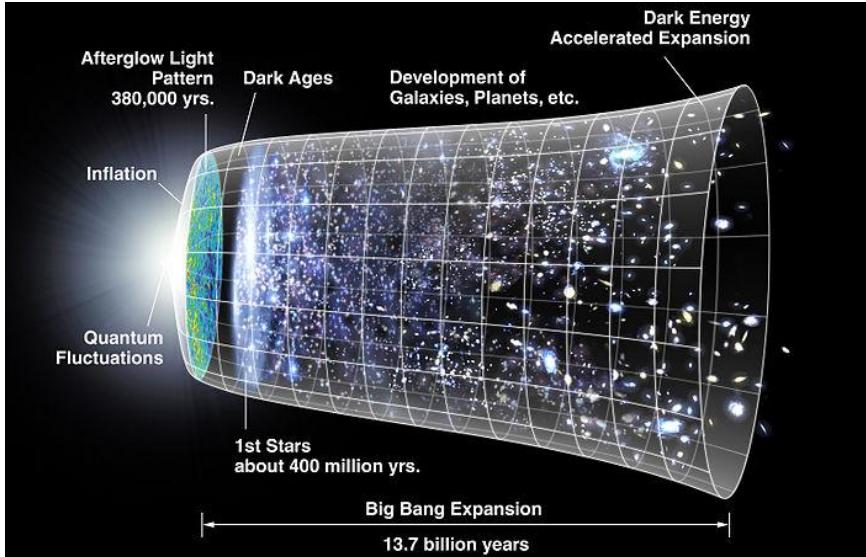
Viernes 12 de Septiembre 2014
6pm-Late

Profesor: Dr. Daniel Almonacid
daniel.almonacid@unab.cl

PARTE 1:

Estructura de Proteínas y

Evolución Molecular

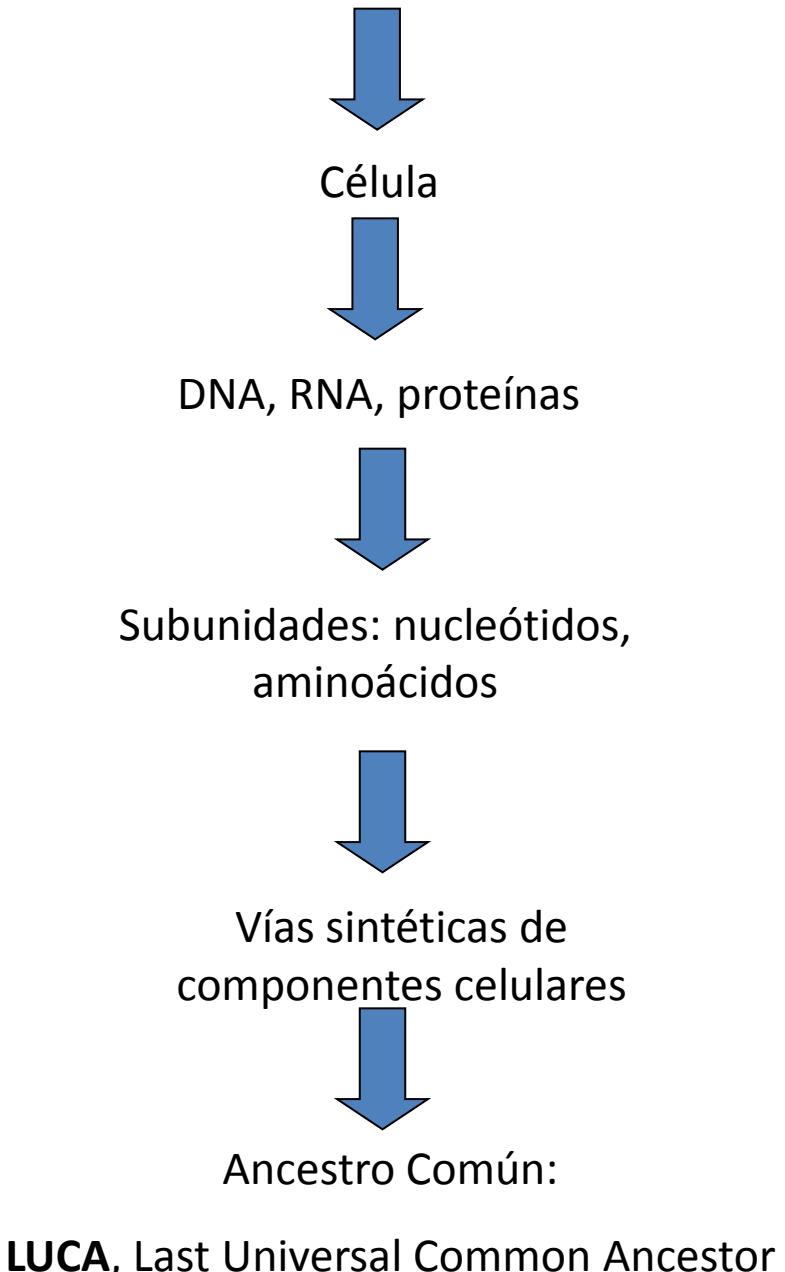


Big Bang ~ 14 mil millones de años

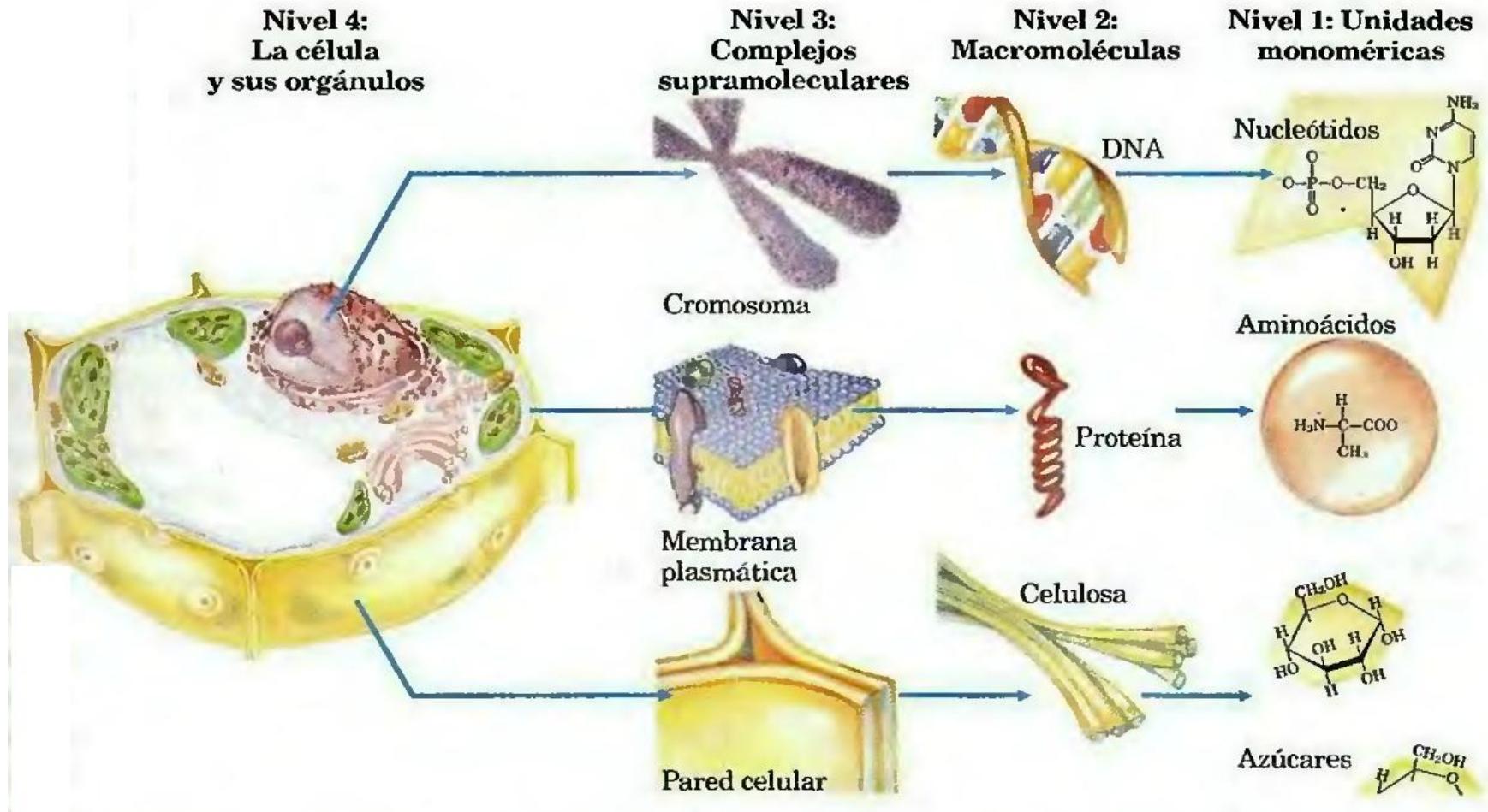


Evolución de Vida en la Tierra
~ 4 mil millones de años

Diversas formas de vida



Jerarquía estructural de la organización molecular de la célula



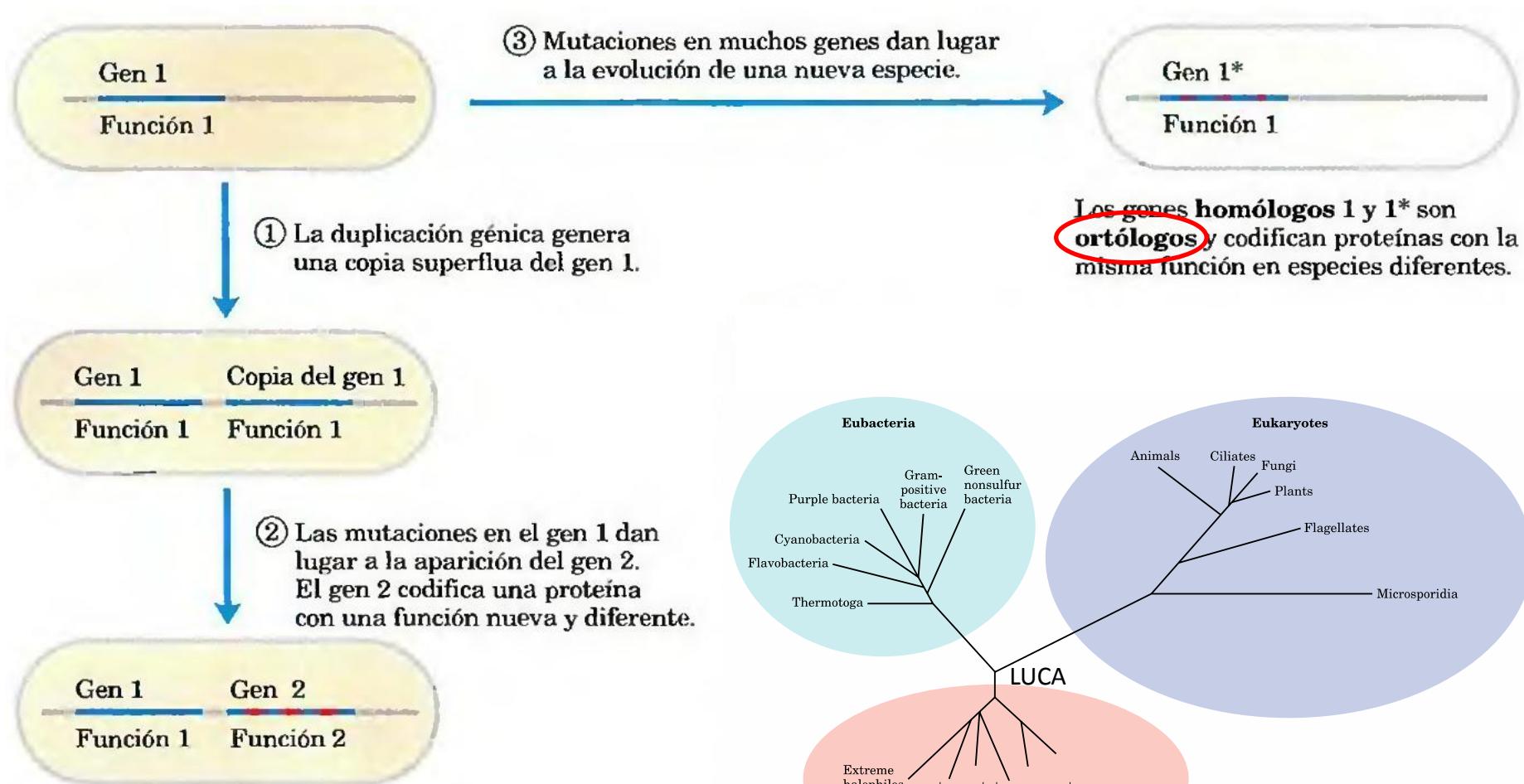
Constitución molecular de una célula

- Las células contienen un conjunto universal de moléculas pequeñas
- Las macromoléculas son las principales constituyentes de las células

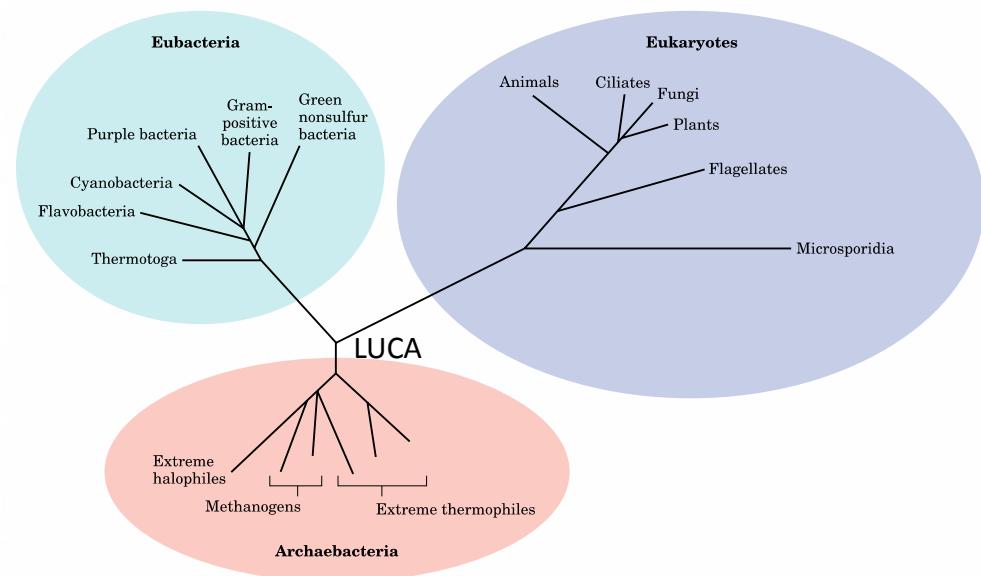
TABLA 1-2 Componentes moleculares de una célula de *E. coli*

	<i>Porcentaje de peso total de la célula</i>	<i>Número aproximado de especies moleculares diferentes</i>
Agua	70	1
Proteínas	15	3000
Ácidos nucleicos		
DNA	1	1
RNA	6	>3000
Polisacáridos	3	5
Lípidos	2	20
Subunidades monoméricas e intermediarios	2	500
Iones inorgánicos	1	20

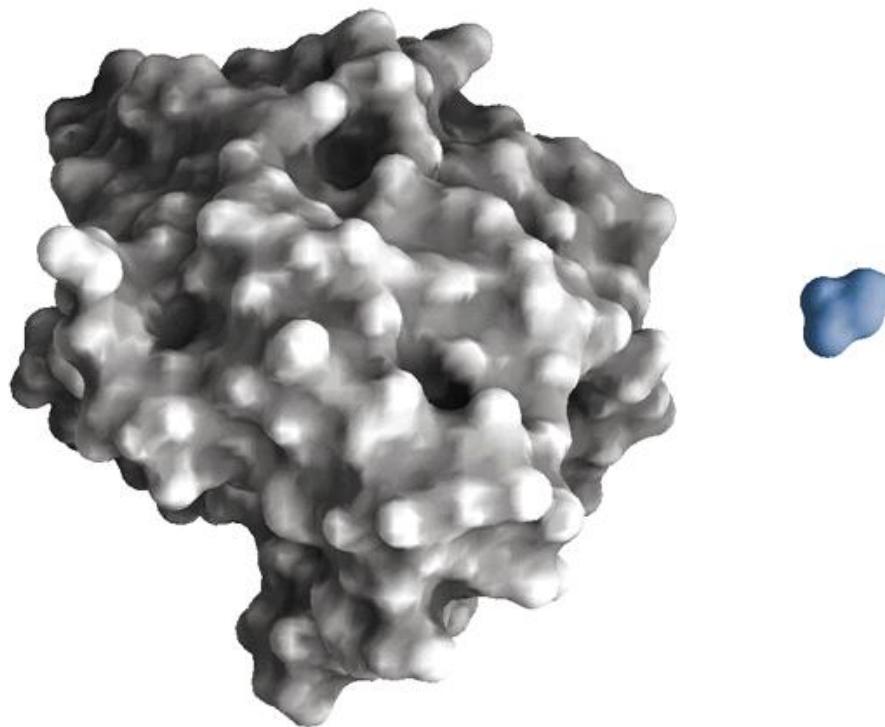
Generación de diversidad genética por mutación y duplicación



Los genes 1 y 2 homólogos son **parálogos**, relacionados secuencialmente pero que codifican proteínas de función diferente dentro de la misma especie.



Estructura = Función de una Proteína



Proteínas

- Son las macromoléculas más abundantes de las células.

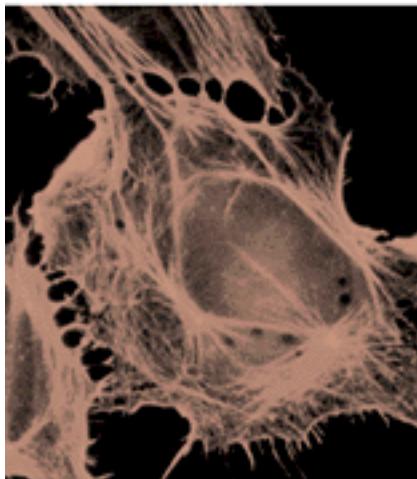
TABLA 1-2 Componentes moleculares de una célula de *E. coli*

	Porcentaje de peso total de la célula	Número aproximado de especies moleculares diferentes
Agua	70	1
Proteínas	15	3000
Ácidos nucleicos		
DNA	1	1
RNA	6	>3000
Polisacáridos	3	5
Lípidos	2	20
Subunidades monoméricas e intermediarios	2	500
Iones inorgánicos	1	20

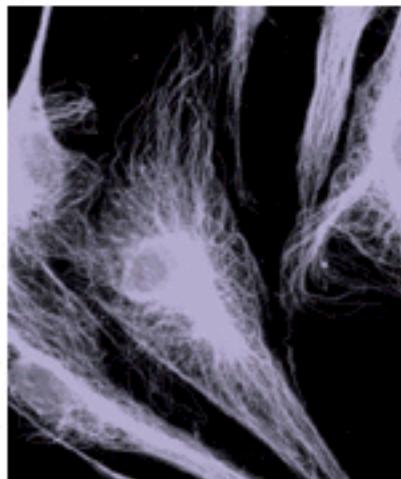
- Cumplen las funciones químicas y biológicas más esenciales en las células

Función Estructural

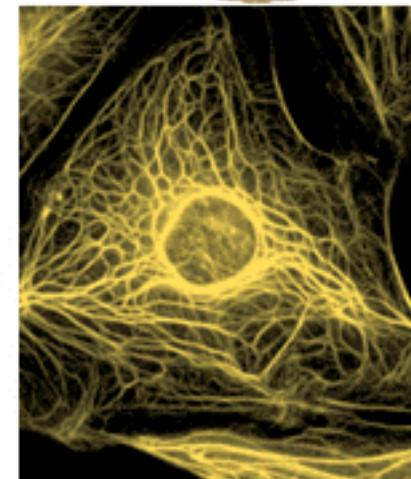
Microfilamentos



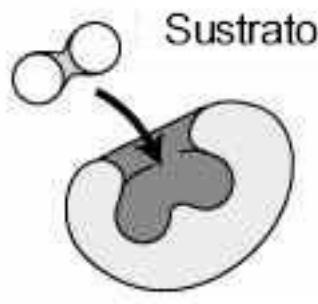
Microtúbulos



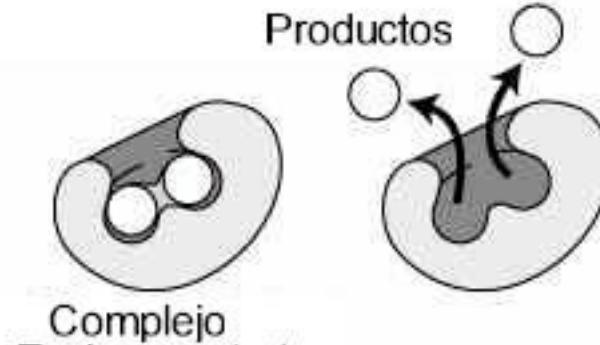
Filamentos intermedios



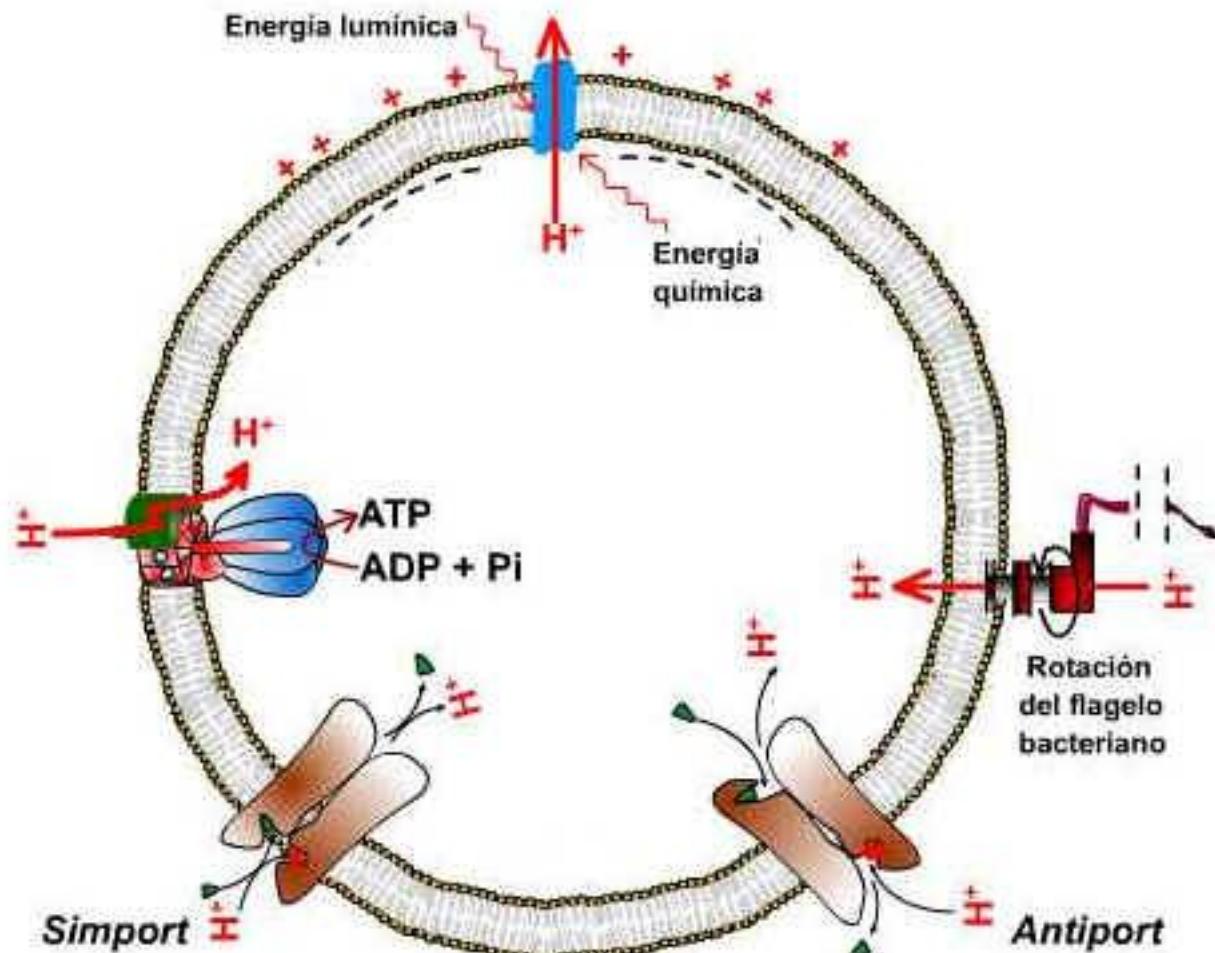
Función Enzimática



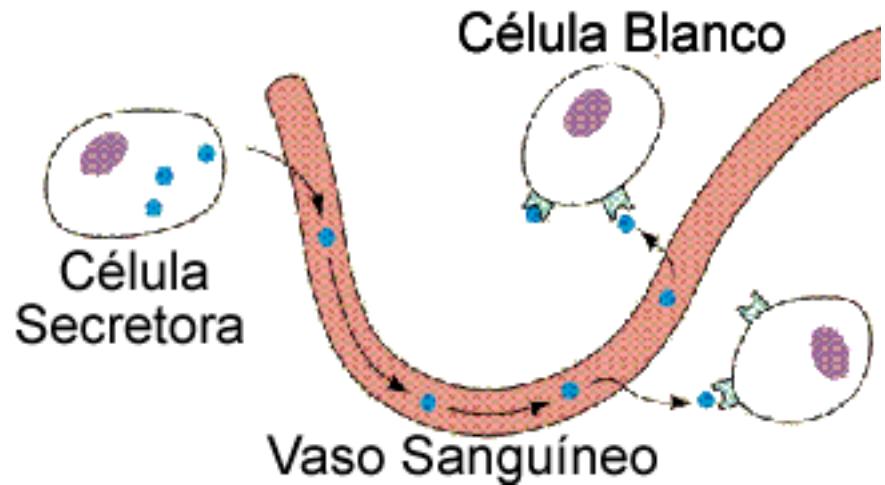
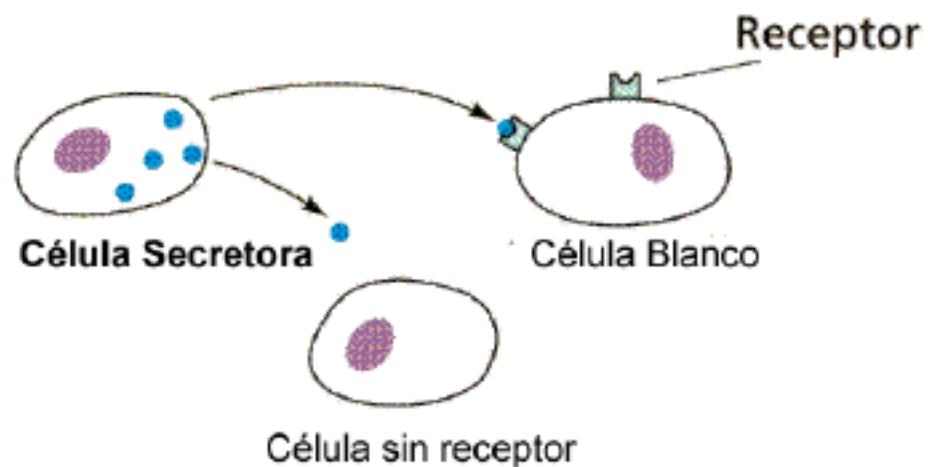
Enzima



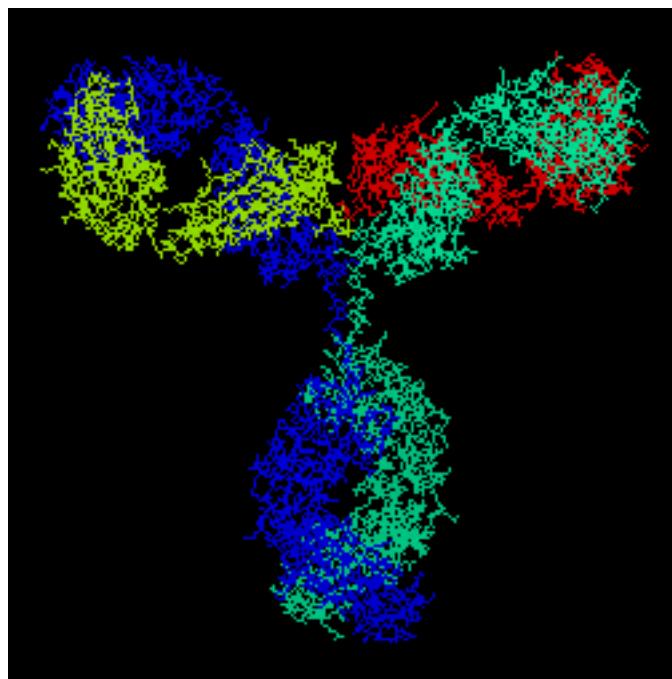
Función de Transporte



Función Hormonal

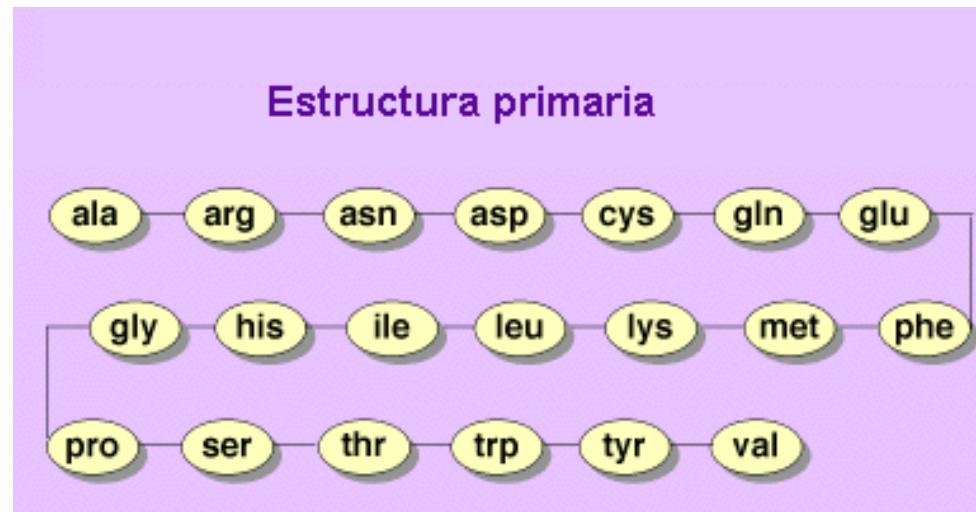


Función de Defensa



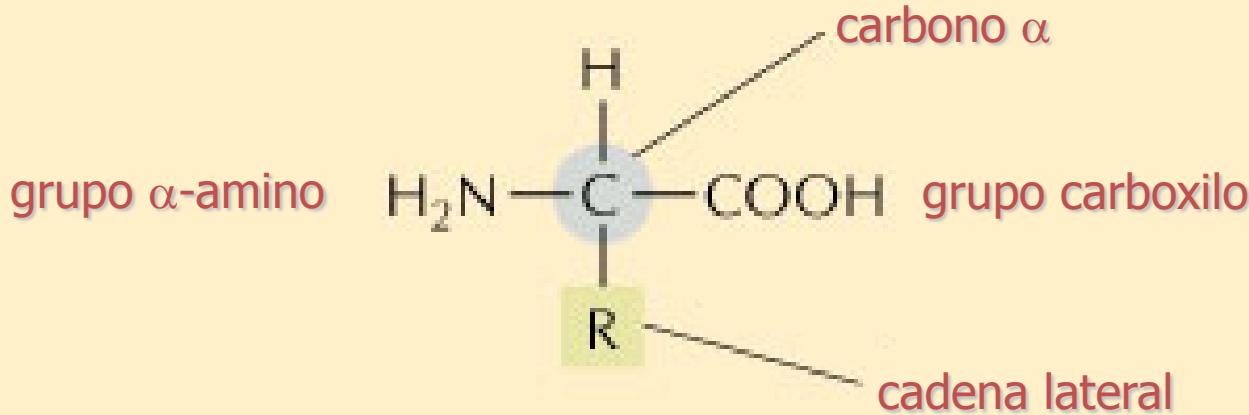
Proteínas

- La función de las proteínas viene estrictamente dada por su estructura.
- Son polímeros biológicos y sus unidades repetitivas son aminoácidos.
- Su tamaño puede variar desde pequeñas moléculas (de dos o tres aminoácidos) hasta macromoléculas de miles de ellos

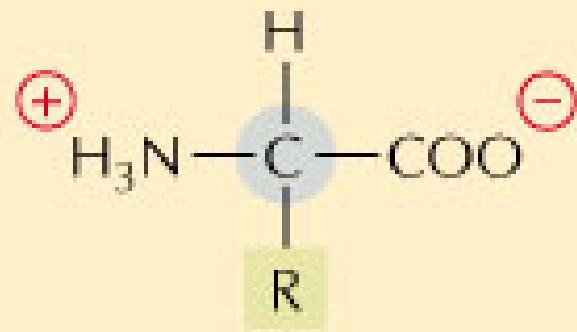


Aminoácidos

La fórmula general de un aminoácido es:

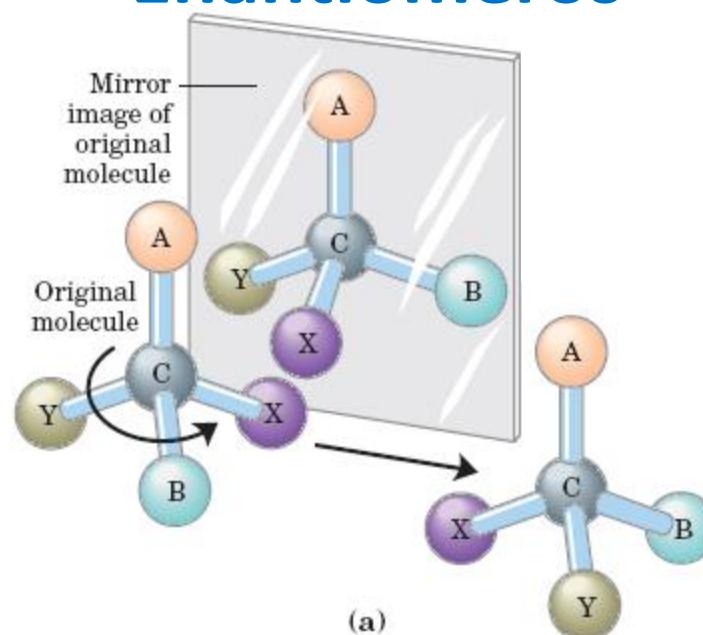


20 tipos diferentes de cadena lateral (R). A pH 7, los grupos amino y carboxilo se encuentran ionizados.

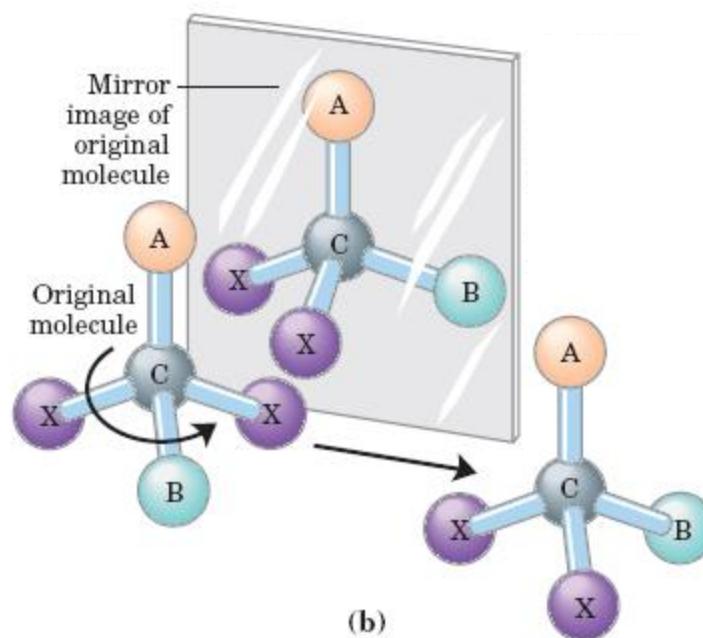


Enantiómeros

Una molécula quiral no puede ser superpuesta sobre su misma imagen al girarla



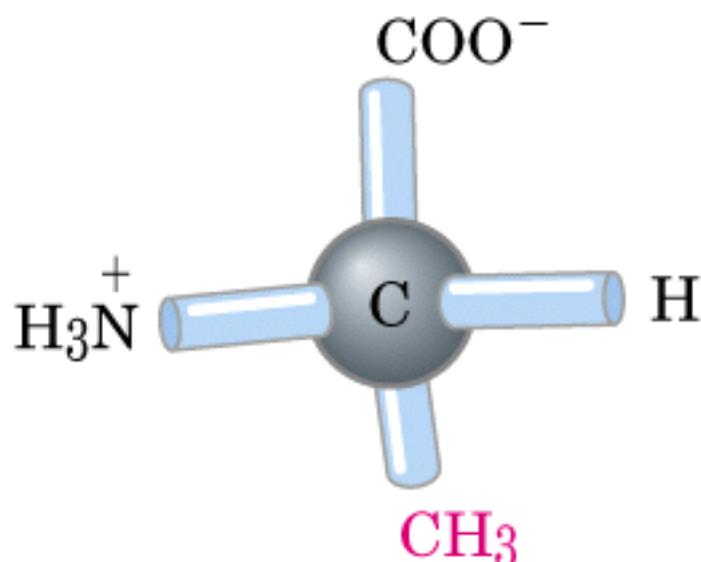
(a)



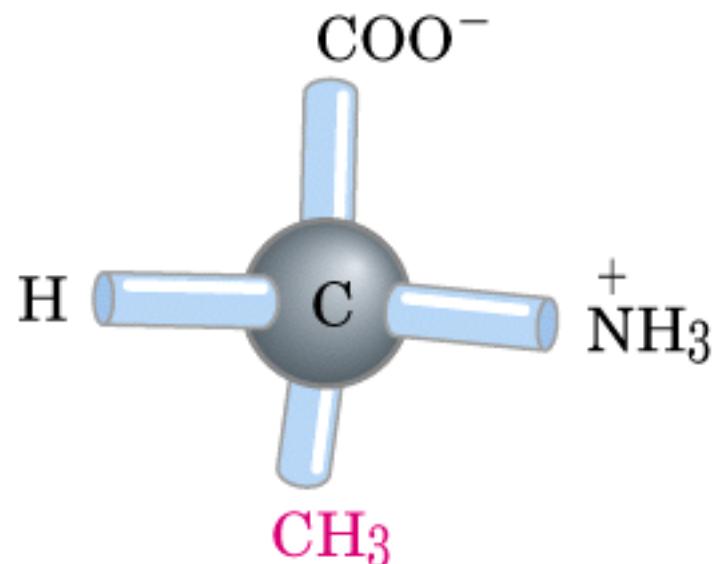
(b)

Una molécula no quiral si puede ser superpuesta sobre su imagen al girarla

Enantiómeros de Alanina



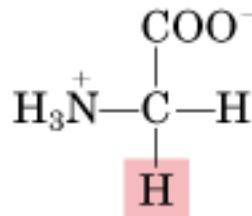
L-Alanine



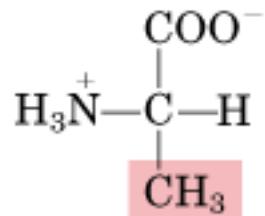
D-Alanine

Clasificación de los Aminoácidos

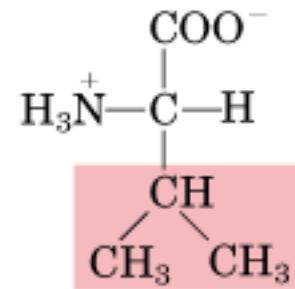
Grupos R alifáticos, no-polares



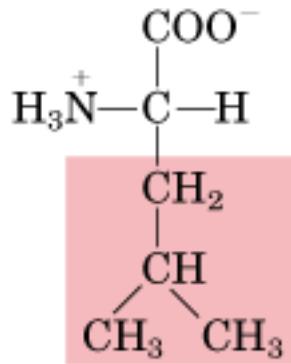
Glycine



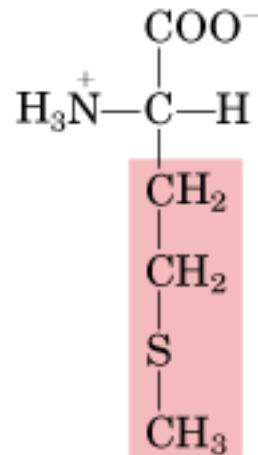
Alanine



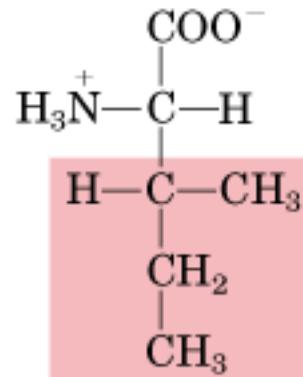
Valine



Leucine

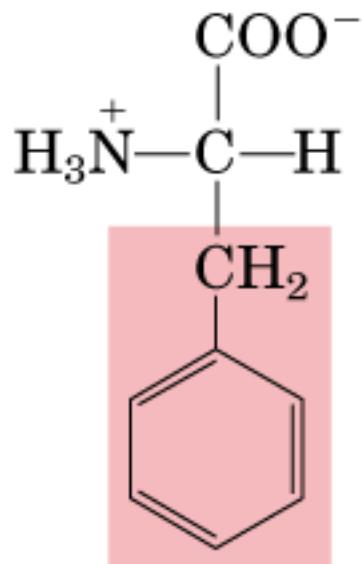


Methionine

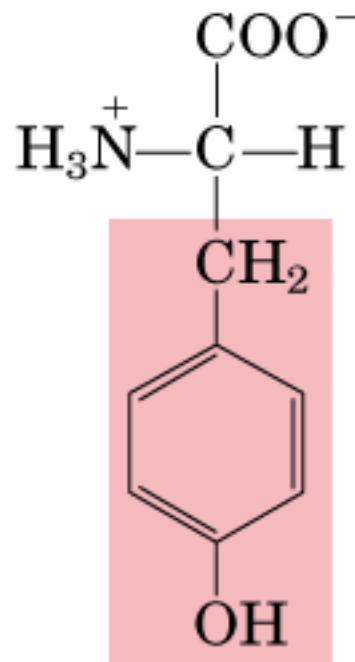


Isoleucine

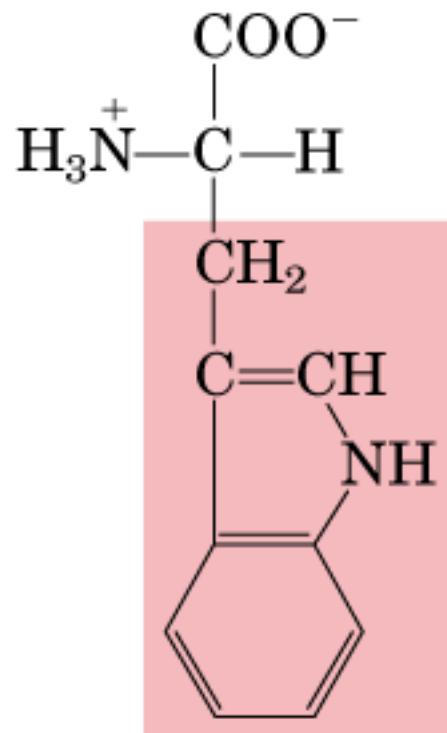
Grupos R aromáticos



Phenylalanine

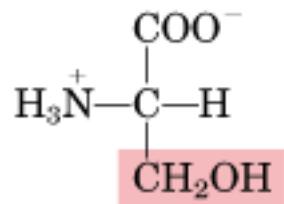


Tyrosine

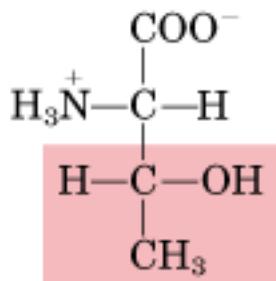


Tryptophan

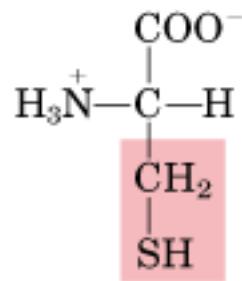
Grupos R sin carga, polares



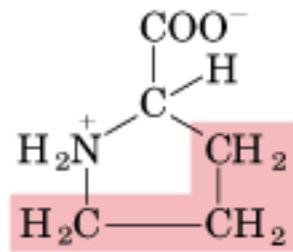
Serine



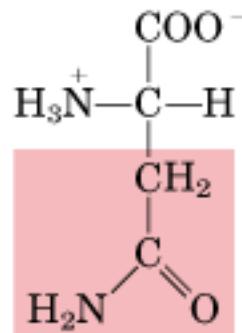
Threonine



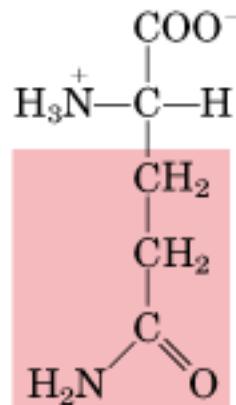
Cysteine



Proline

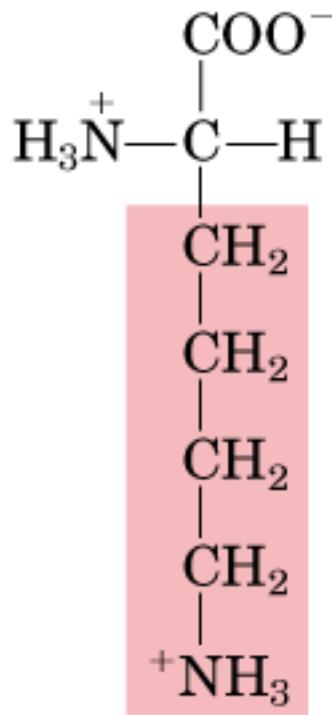


Asparagine

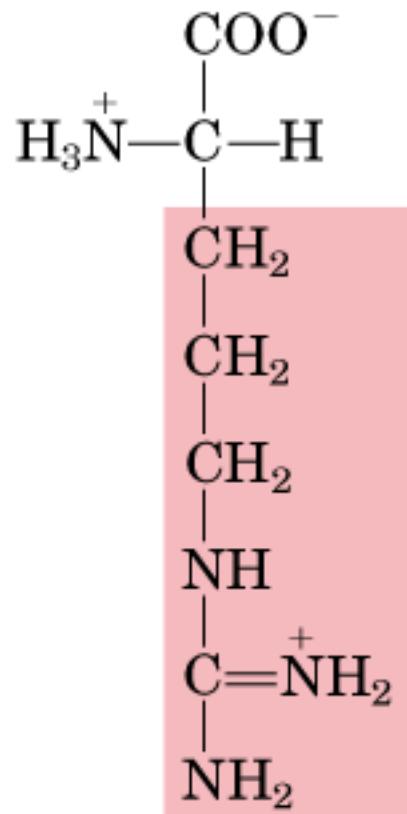


Glutamine

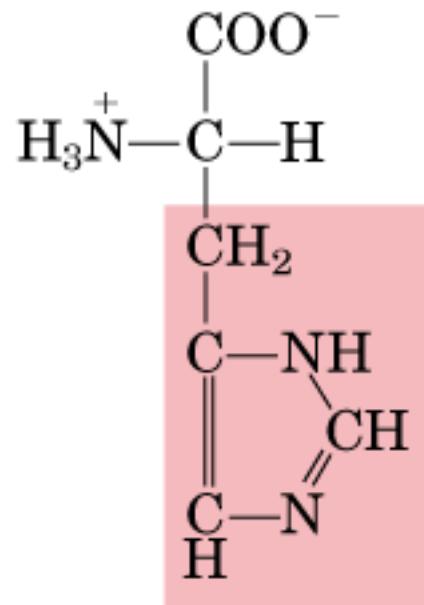
Grupos R con carga positiva



Lysine

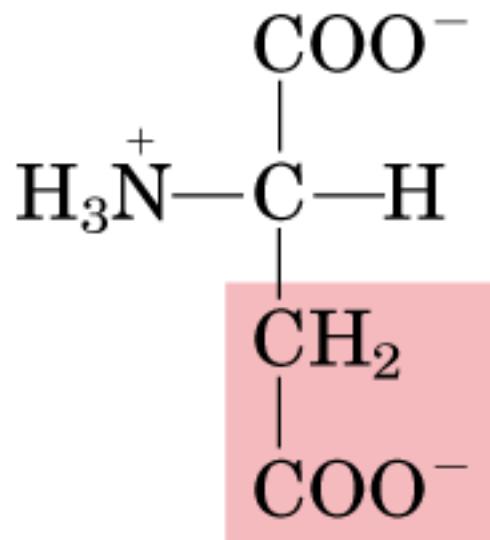


Arginine

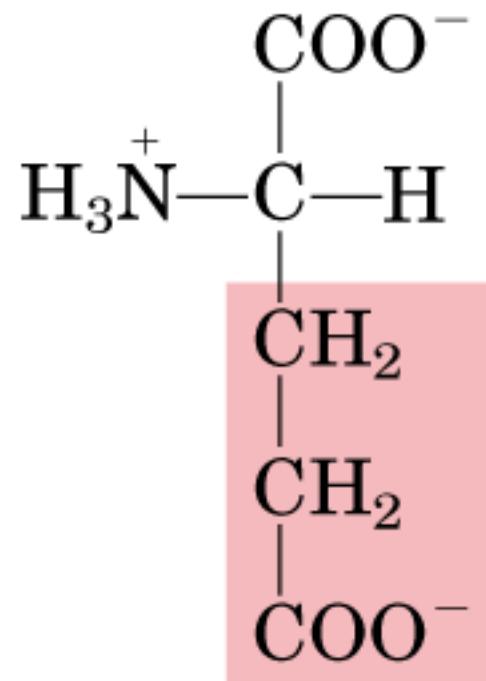


Histidine

Grupos R con carga negativa



Aspartate



Glutamate

table 5–1

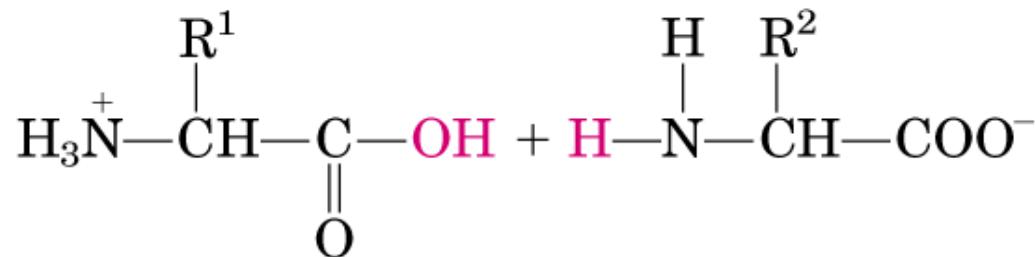
Properties and Conventions Associated with the Standard Amino Acids

Amino acid	Abbreviated names	M_r	pK_a values			pl	Hydropathy index*	Occurrence in proteins (%)†
			pK_1 (—COOH)	pK_2 (—NH ₃ ⁺)	pK_R (R group)			
Nonpolar, aliphatic R groups								
Glycine	Gly G	75	2.34	9.60		5.97	-0.4	7.2
Alanine	Ala A	89	2.34	9.69		6.01	1.8	7.8
Valine	Val V	117	2.32	9.62		5.97	4.2	6.6
Leucine	Leu L	131	2.36	9.60		5.98	3.8	9.1
Isoleucine	Ile I	131	2.36	9.68		6.02	4.5	5.3
Methionine	Met M	149	2.28	9.21		5.74	1.9	2.3
Aromatic R groups								
Phenylalanine	Phe F	165	1.83	9.13		5.48	2.8	3.9
Tyrosine	Tyr Y	181	2.20	9.11	10.07	5.66	-1.3	3.2
Tryptophan	Trp W	204	2.38	9.39		5.89	-0.9	1.4
Polar, uncharged R groups								
Serine	Ser S	105	2.21	9.15		5.68	-0.8	6.8
Proline	Pro P	115	1.99	10.96		6.48	1.6	5.2
Threonine	Thr T	119	2.11	9.62		5.87	-0.7	5.9
Cysteine	Cys C	121	1.96	10.28	8.18	5.07	2.5	1.9
Asparagine	Asn N	132	2.02	8.80		5.41	-3.5	4.3
Glutamine	Gln Q	146	2.17	9.13		5.65	-3.5	4.2
Positively charged R groups								
Lysine	Lys K	146	2.18	8.95	10.53	9.74	-3.9	5.9
Histidine	His H	155	1.82	9.17	6.00	7.59	-3.2	2.3
Arginine	Arg R	174	2.17	9.04	12.48	10.76	-4.5	5.1
Negatively charged R groups								
Aspartate	Asp D	133	1.88	9.60	3.65	2.77	-3.5	5.3
Glutamate	Glu E	147	2.19	9.67	4.25	3.22	-3.5	6.3

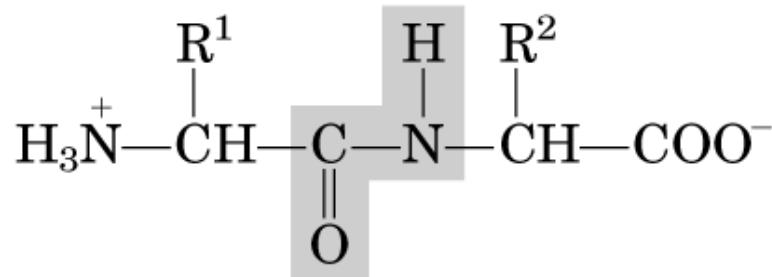
*A scale combining hydrophobicity and hydrophilicity of R groups; it can be used to measure the tendency of an amino acid to seek an aqueous environment (− values) or a hydrophobic environment (+ values). See Chapter 12. From Kyte, J. & Doolittle, R.F. (1982) *J. Mol. Biol.* **157**, 105–132.

†Average occurrence in over 1150 proteins. From Doolittle, R.F. (1989) Redundancies in protein sequences. In *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G.D., ed) Plenum Press, NY, pp. 599–623.

Enlace Peptídico

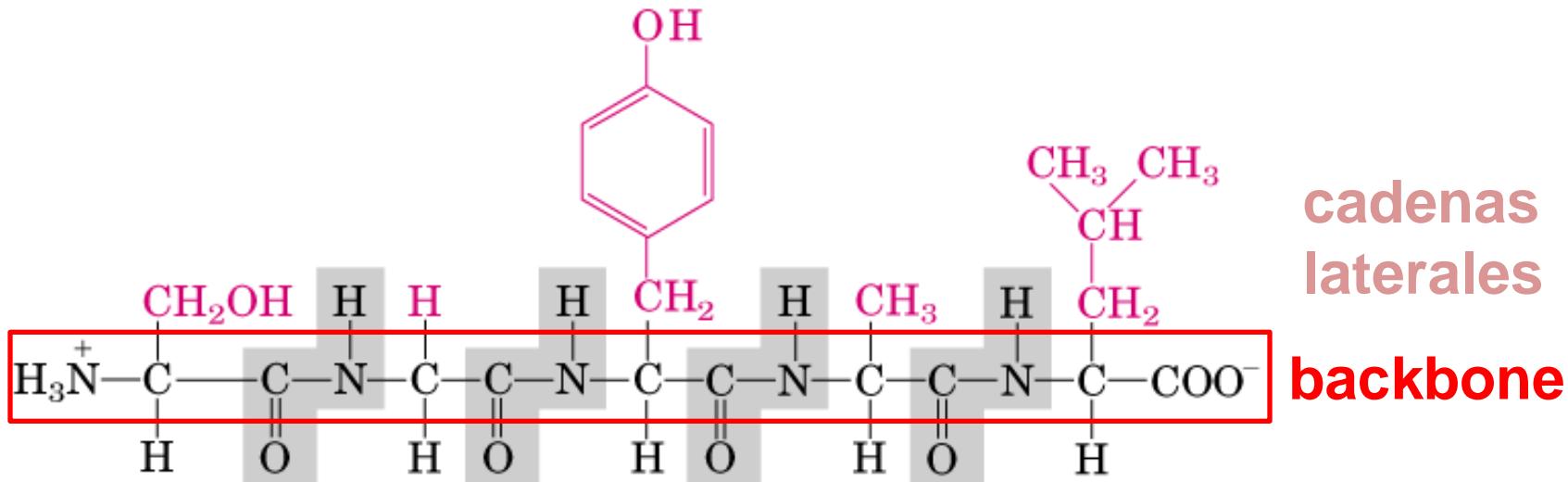


CONDENSACION
(deshidratación)



DIPEPTIDO (dos residuos de aminoácidos)

Enlace Peptídico



Amino terminal → Carboxilo terminal

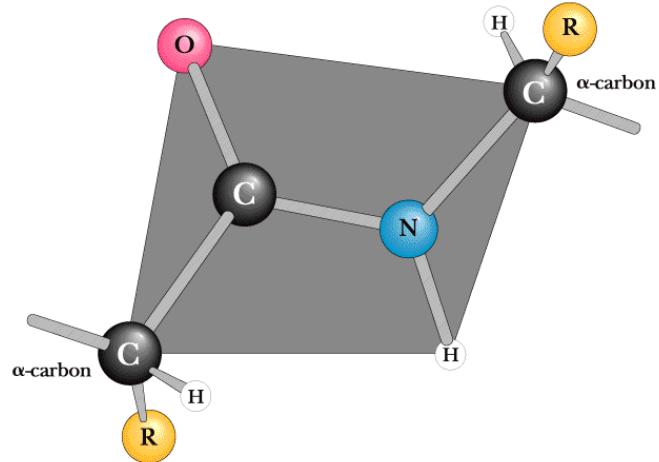
Ser – Gly – Tyr – Ala – Leu

1 – 2 – 3 – 4 – 5

< 50 residuos: **péptido**. No tiene estructura 3D bien definida

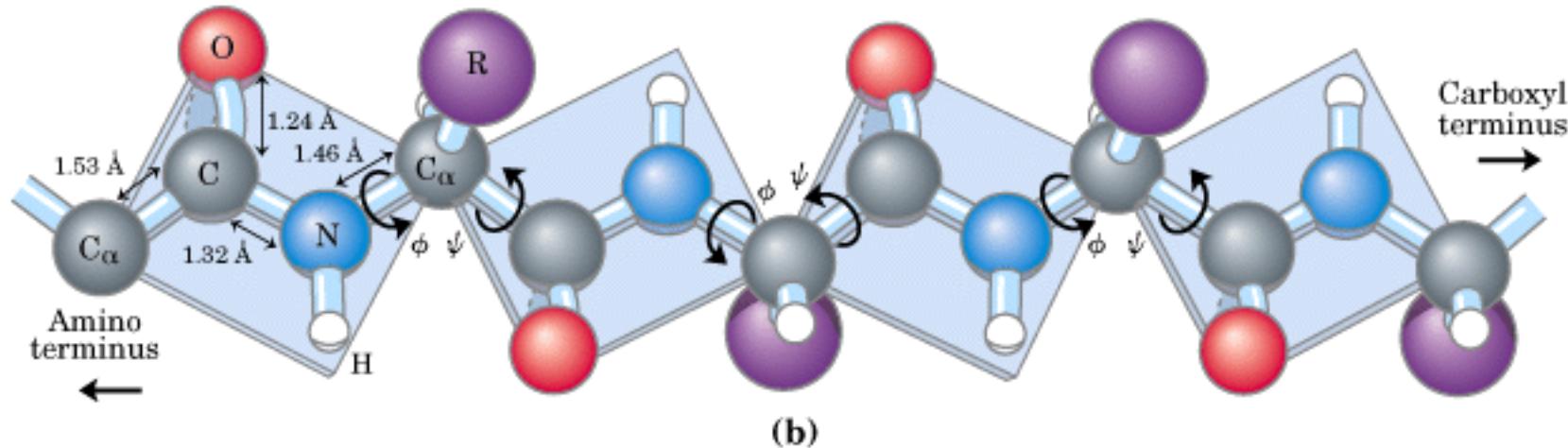
> 50 residuos: **proteína**. Tiene estructura 3D bien definida

Carácter del Enlace Peptídico

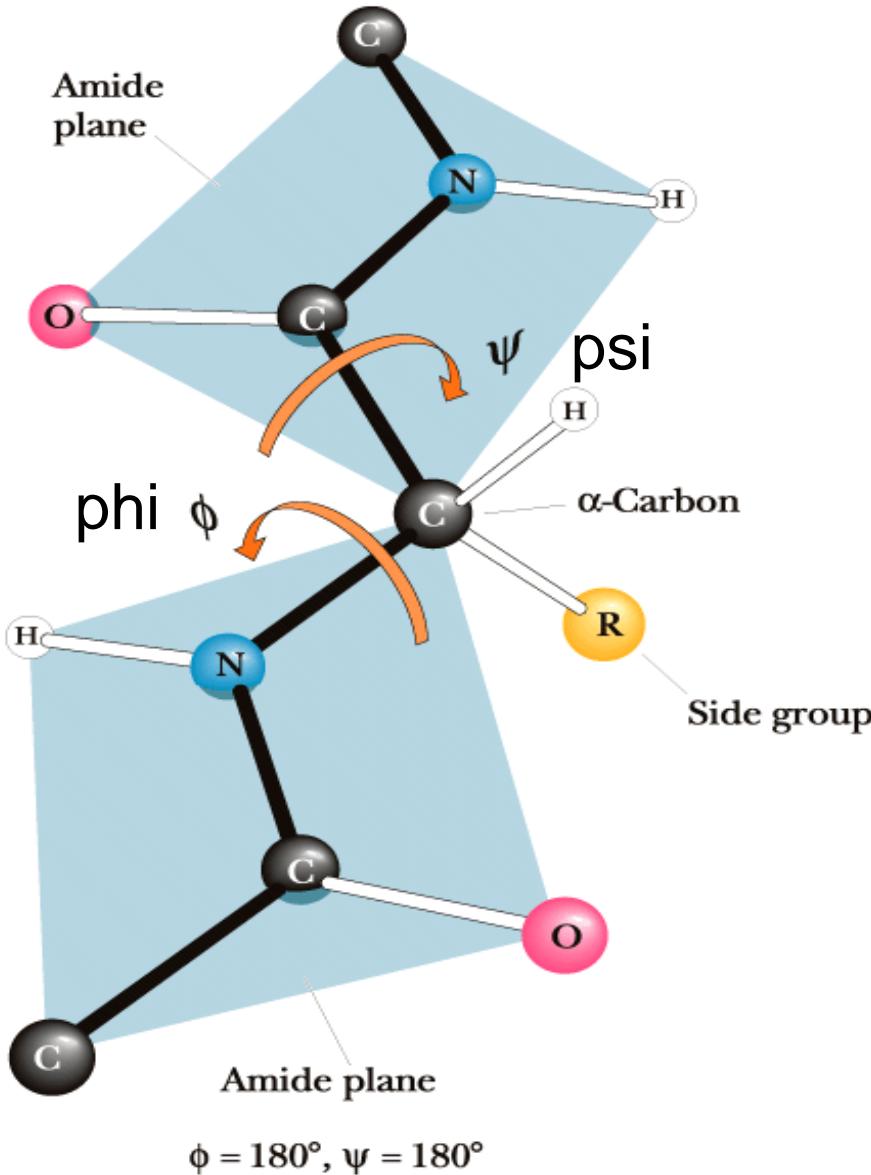


El enlace peptídico C-N es parcialmente un doble enlace, y por tanto es **rígido** y **plano**

Existen ángulos de rotación entre planos, llamados **ángulos dihedros**.



Angulos Dihedros



Distintas combinaciones de ángulos phi y psi permiten que las proteínas se empaqueten en diferentes formas.

Gráfico de Ramachandran

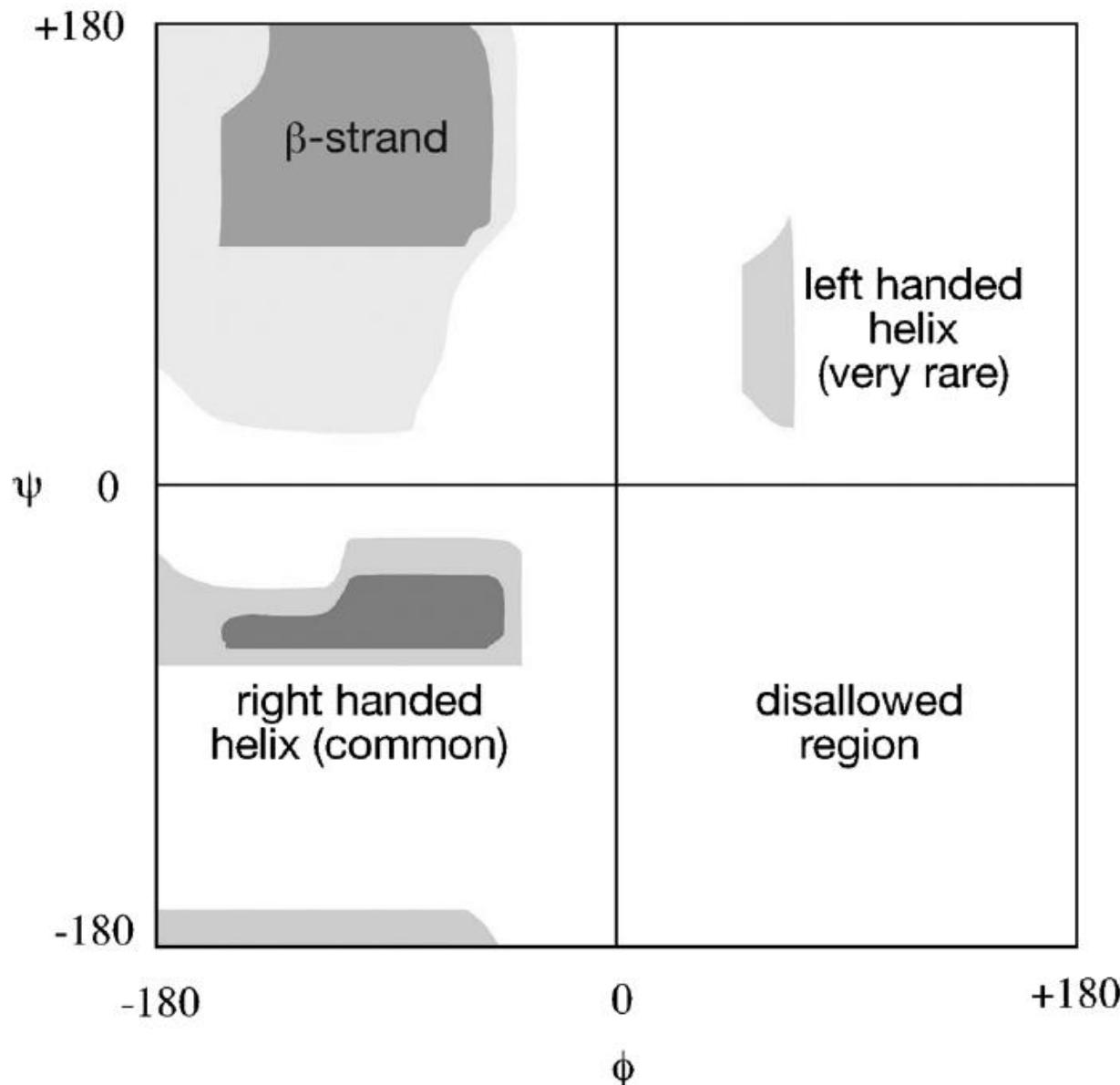
Gráfico que **muestra los ángulos φ y ψ para una proteína completa**, dando cuenta del **espacio conformacional de la proteína**.

La **rotación** de los ángulos φ y ψ **no es completamente libre** debido a la **naturaleza plana del enlace peptídico** y el **impedimento estérico** que ejercen las cadenas laterales.

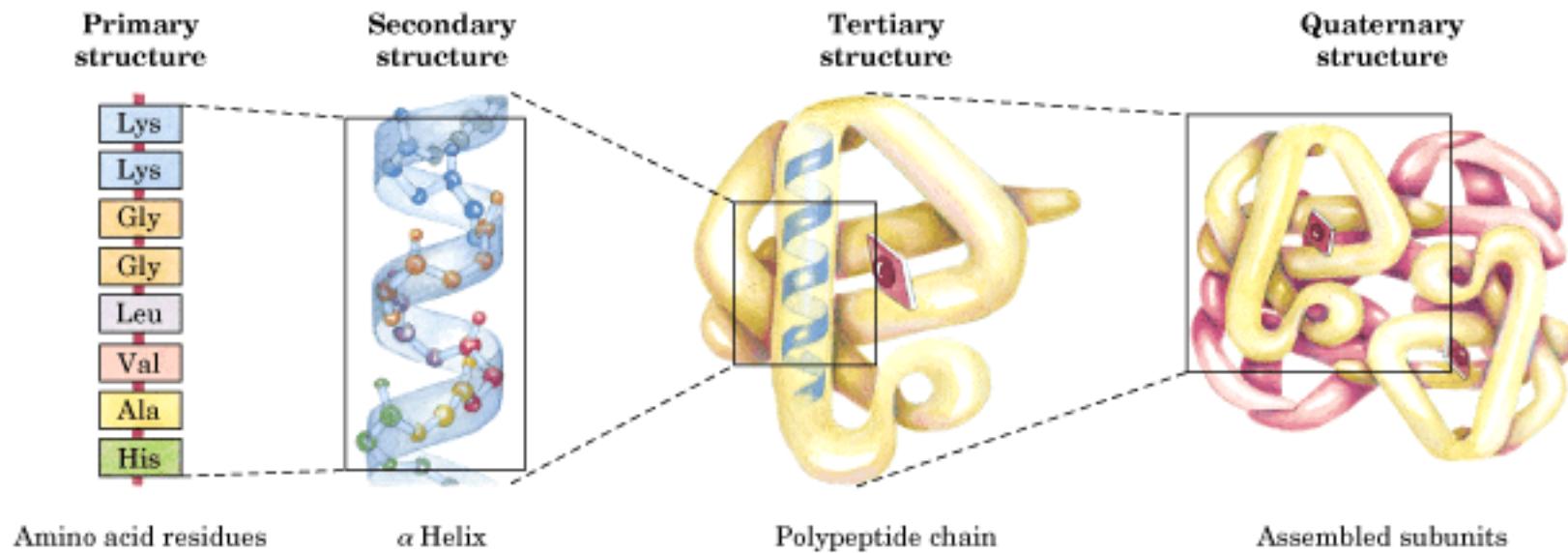
Permite identificar **regiones de conformación que están estericamente permitidas y otras no permitidas**.

Permite evaluar la calidad de modelos de proteínas.

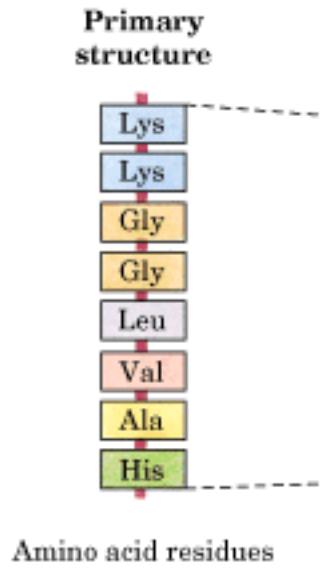
Gráfico de Ramachandran



Estructura de Proteínas



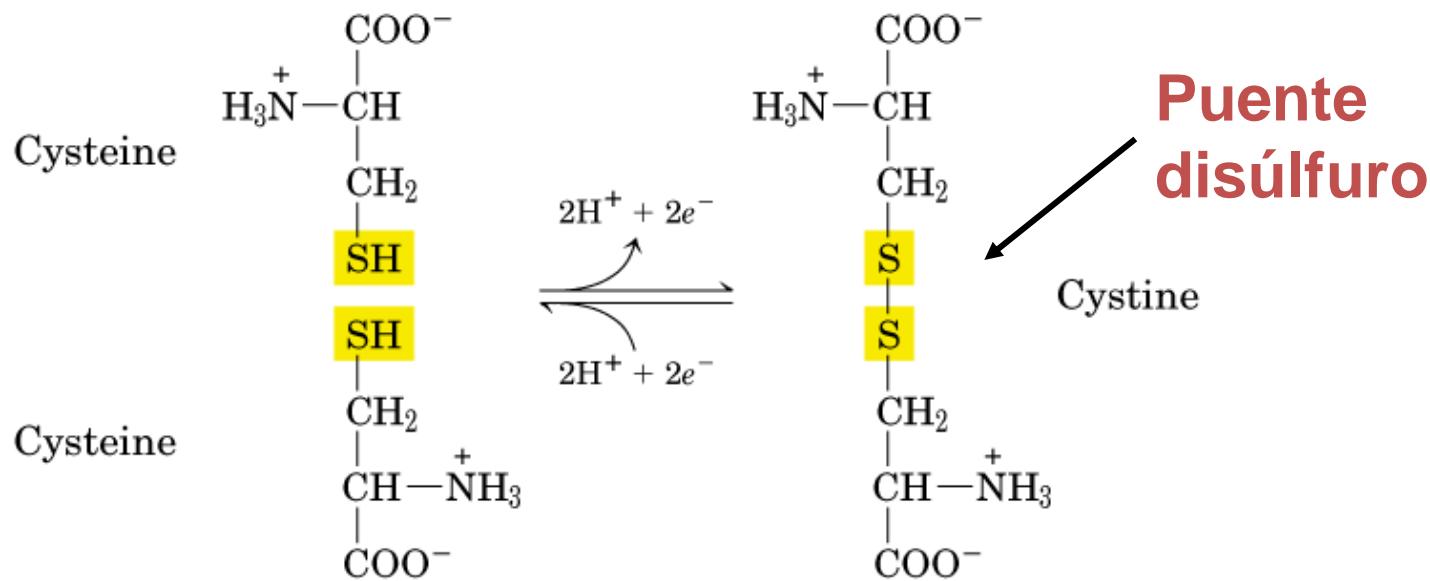
Estructura Primaria



La **estructura primaria** corresponde a la **secuencia lineal de aminoácidos** unidos por **enlaces peptídicos** y también por **puentes disulfuro**

Estructura Primaria: Multi-subunidades

Son proteínas que poseen **dos o más cadenas polipeptídicas asociadas por enlaces covalentes disulfuro.**



Ejemplo: Estructura primaria de la insulina

A chain		B chain
NH_3^+		NH_3^+
Gly		Phe
Ile		Val
Val		Asn
Gln		Gln
5 Gln		5 His
Cys		Leu
Cys	—S—S—Cys	
Ala		Gly
Ser		Ser
10 Val		10 His
Cys		Leu
Ser		Val
Leu		Glu
Tyr		Ala
15 Gln		15 Leu
Leu		Tyr
Glu		Leu
Asn		Val
Tyr		
20 Cys	—S—Cys	
Asn		Glu
COO ⁻		Arg
		Gly
		Phe
		25 Phe
		Tyr
		Thr
		Pro
		Lys
		30 Ala
		COO ⁻

Fuerzas que Estabilizan las Estructuras 2^a a 4^a de Proteínas

table 4–4

Four Types of Noncovalent ("Weak") Interactions among Biomolecules in Aqueous Solvent

Hydrogen bonds

Between neutral groups



Between peptide bonds



Ionic interactions

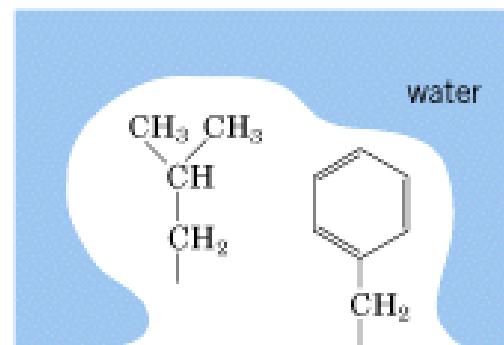
Attraction



Repulsion



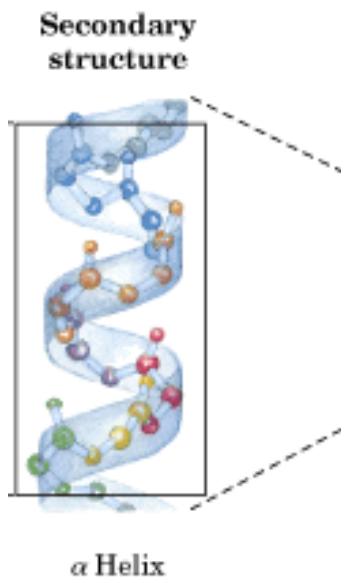
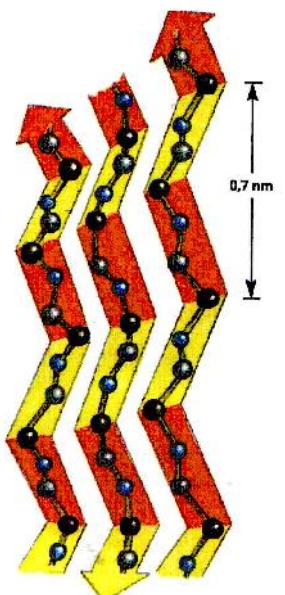
Hydrophobic interactions



Van der Waals interactions

Any two atoms in close proximity

Estructura Secundaria

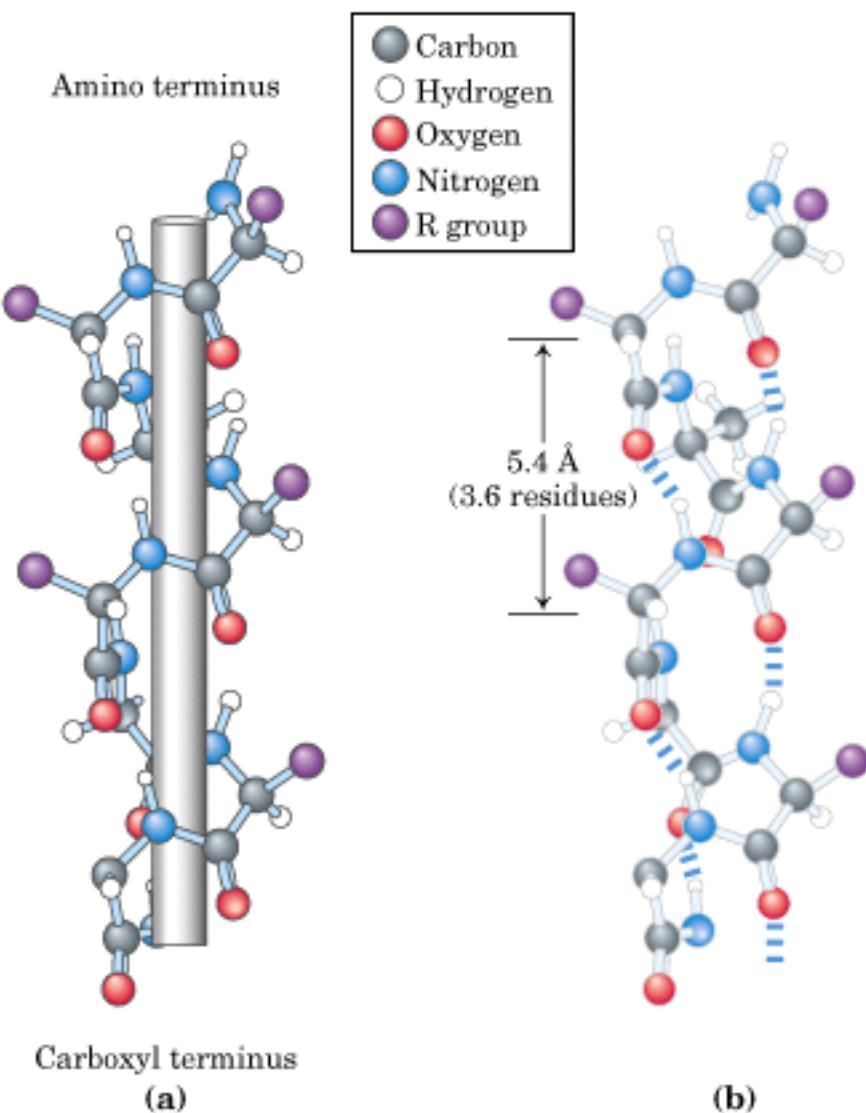


La **estructura secundaria** corresponde a una **conformación local estable de aminoácidos**.

Se caracteriza por el **arreglo regular y repetitivo de residuos** de aminoácidos.

Se **estabiliza** gracias a los **puentes de hidrógeno entre** los átomos del “main chain” del **grupo C=O y el grupo N-H** de diferentes residuos.

Estructura Secundaria: α -hélice



La mayoría de las α -hélice **dan vuelta a la derecha**, y tienen **3.6 residuos por vuelta**.

La estructura se estabiliza por **puentes de hidrógeno entre residuos i e $i+4$** .

Los ángulos φ y ψ promedio son -60 y -45, respectivamente (**región inferior izquierda en el gráfico de Ramachandran**).

Residuos hidrofóbicos están hacia la cara interna de la hélice, y los hidrofílicos hacia la externa.

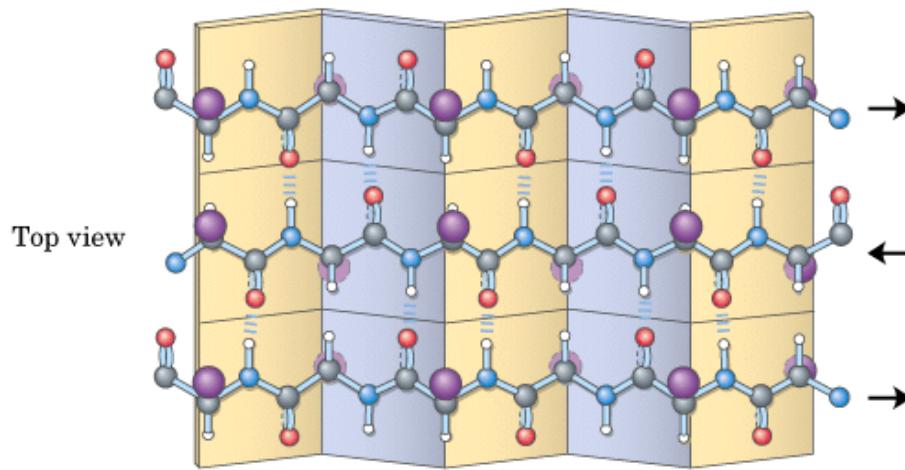
Estructura Secundaria: sábana- β

Son **configuraciones completamente extendidas**.

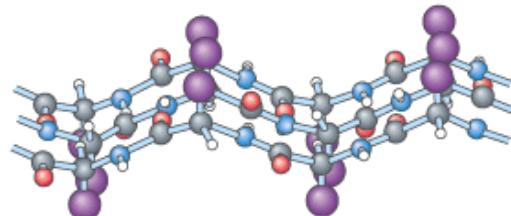
La estructura se estabiliza por **puentes de hidrógeno entre residuos de sábanas adyacentes**.

Los ángulos φ y ψ caen en la **región superior izquierda en el gráfico de Ramachandran**.

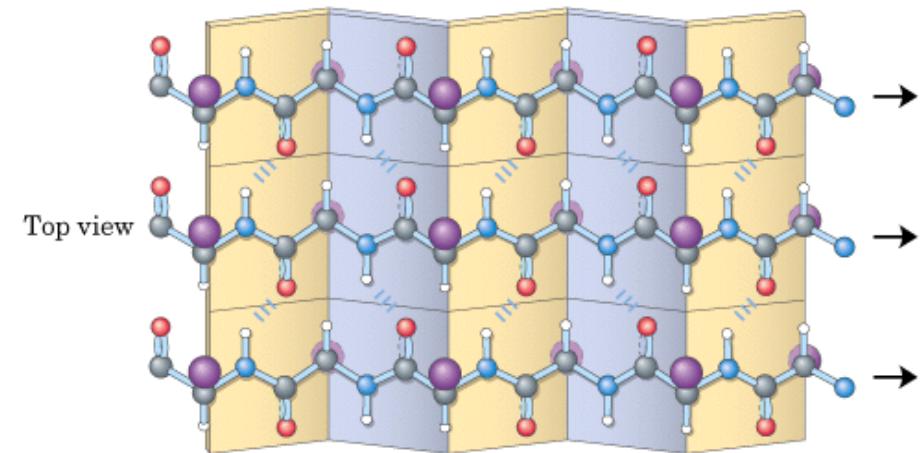
(a) Antiparallel



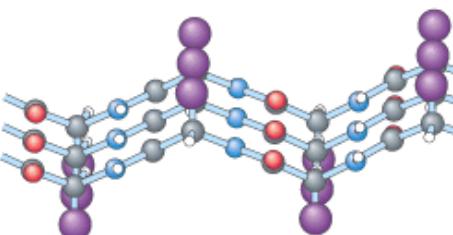
Side view



(b) Parallel



Side view



Estructura Secundaria: coils y loops

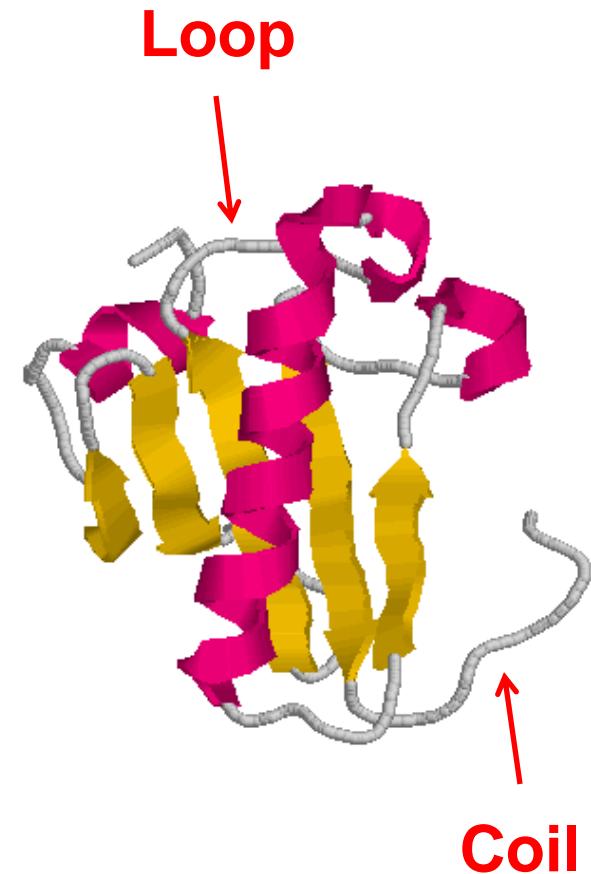
Estas son **estructuras irregulares**.

Loops se caracterizan por tener vueltas muy marcadas o estructuras tipo “pinche”.

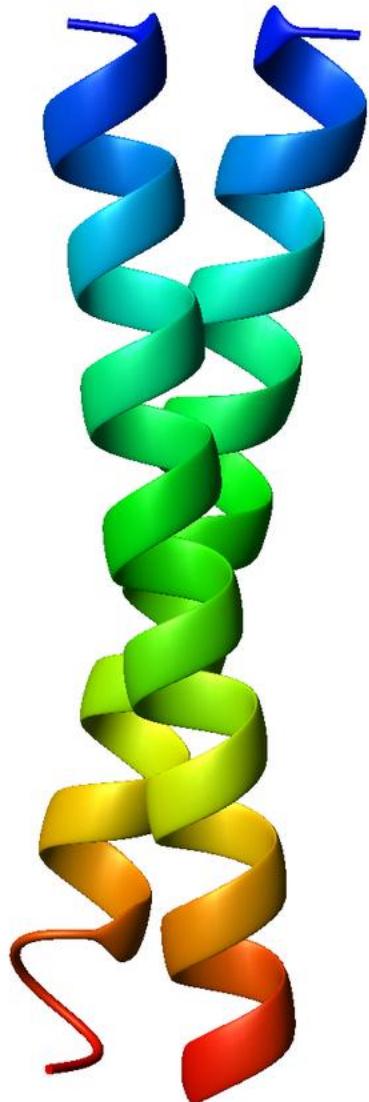
Si las regiones que rodean la estructura son también irregulares, entonces reciben el nombre de coils.

Los residuos en los loops y coils son generalmente cargados y polares y se localizan en la superficie de las proteínas.

En términos evolutivos, estas regiones son variables, donde frecuentemente ocurren mutaciones, delecciones e inserciones.



Estructura Secundaria: coiled coils



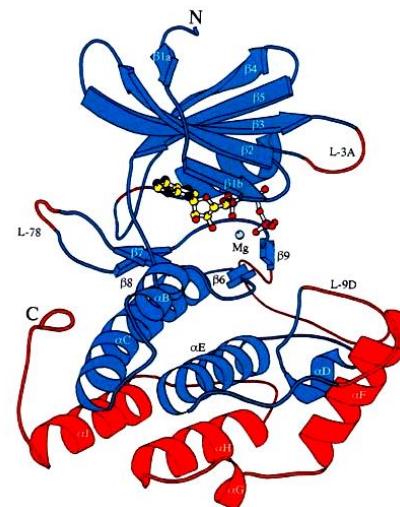
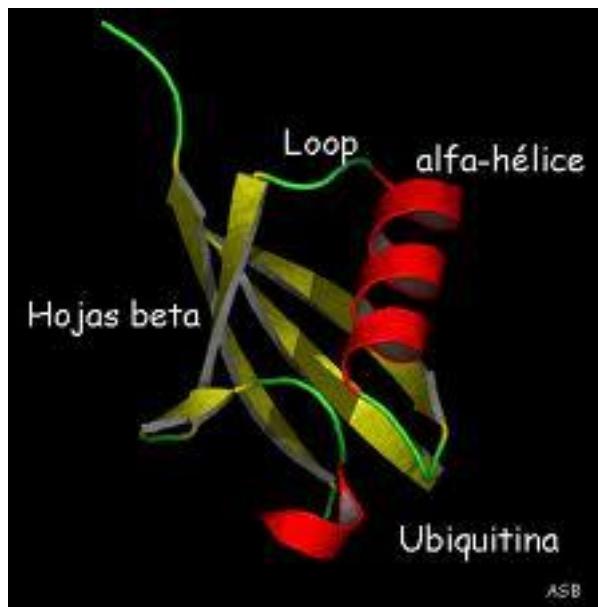
Coiled coils se caracterizan por formar **paquetes de dos o más α -helices.**

Las hélices que forman los coiled coils tienen un **patrón especial de hidrofobicidad que se repite cada 7 residuos**: 5 residuos polares y 2 hidrófobicos: HPPHPPP

Estructura Terciaria

La estructura terciaria corresponde al **plegamiento en tres dimensiones de toda la cadena polipeptídica**.

Incluye varios elementos de estructura secundaria y sus regiones circundantes.

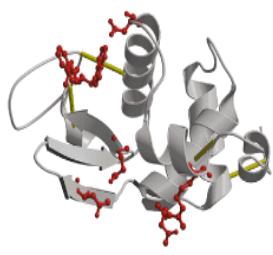
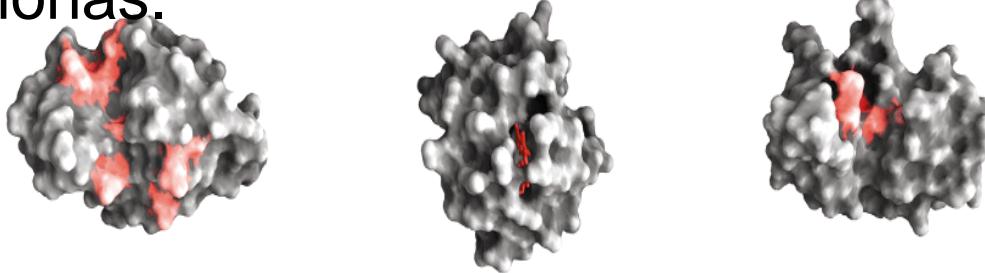


Estructura Terciaria: Proteínas Globulares

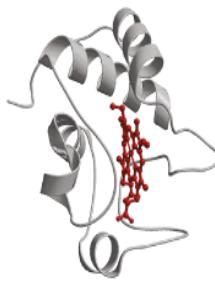
Son **solubles** y rodeadas de moléculas de agua.

Tienen estructuras **compactas o esféricas con residuos polares o hidrofílicos en la superficie y residuos hidrofóbicos hacia el centro** de la proteína.

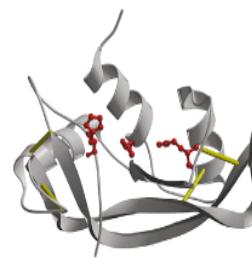
Ejemplos de proteínas globulares : enzimas, mioglobinas, citokinas, hormonas.



Lysozyme



Cytochrome c



Ribonuclease

Estructura Terciaria: Proteínas de Membrana

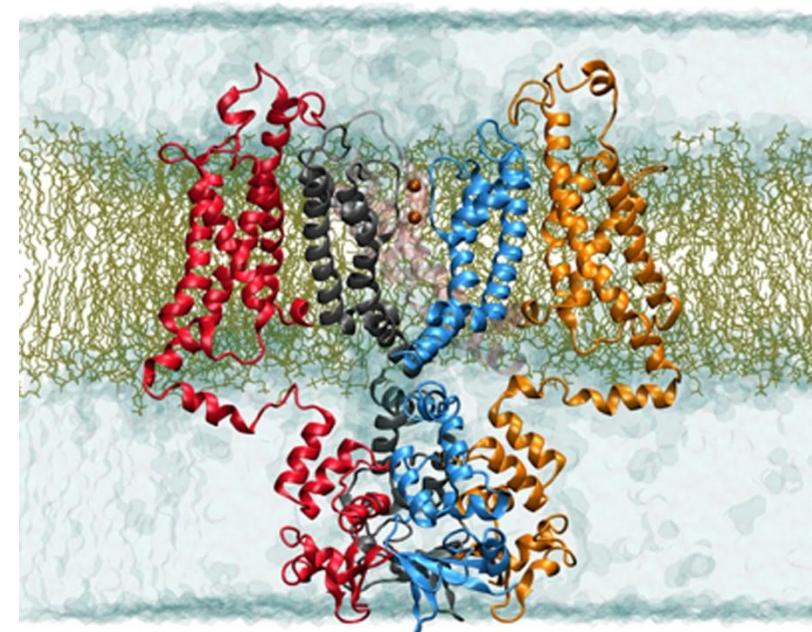
Las proteínas de membrana
**existen en las bicapas lipídicas
de las membranas celulares.**

Debido a que estan rodeadas de
lípidos, **su exterior es muy
hidrofóbico.**

**La mayoría de los segmentos
transmembrana son α -hélices.**

Sin embargo, existen proteínas de
membrana que están compuestas
de **sábanas- β .**

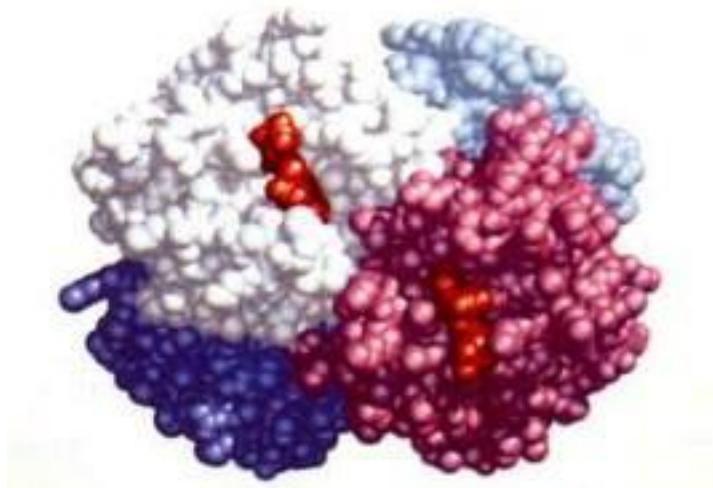
Ejemplos de proteínas de
membrana son: rhodopsinas,
citocromo c oxidasa, y canales
iónicos.



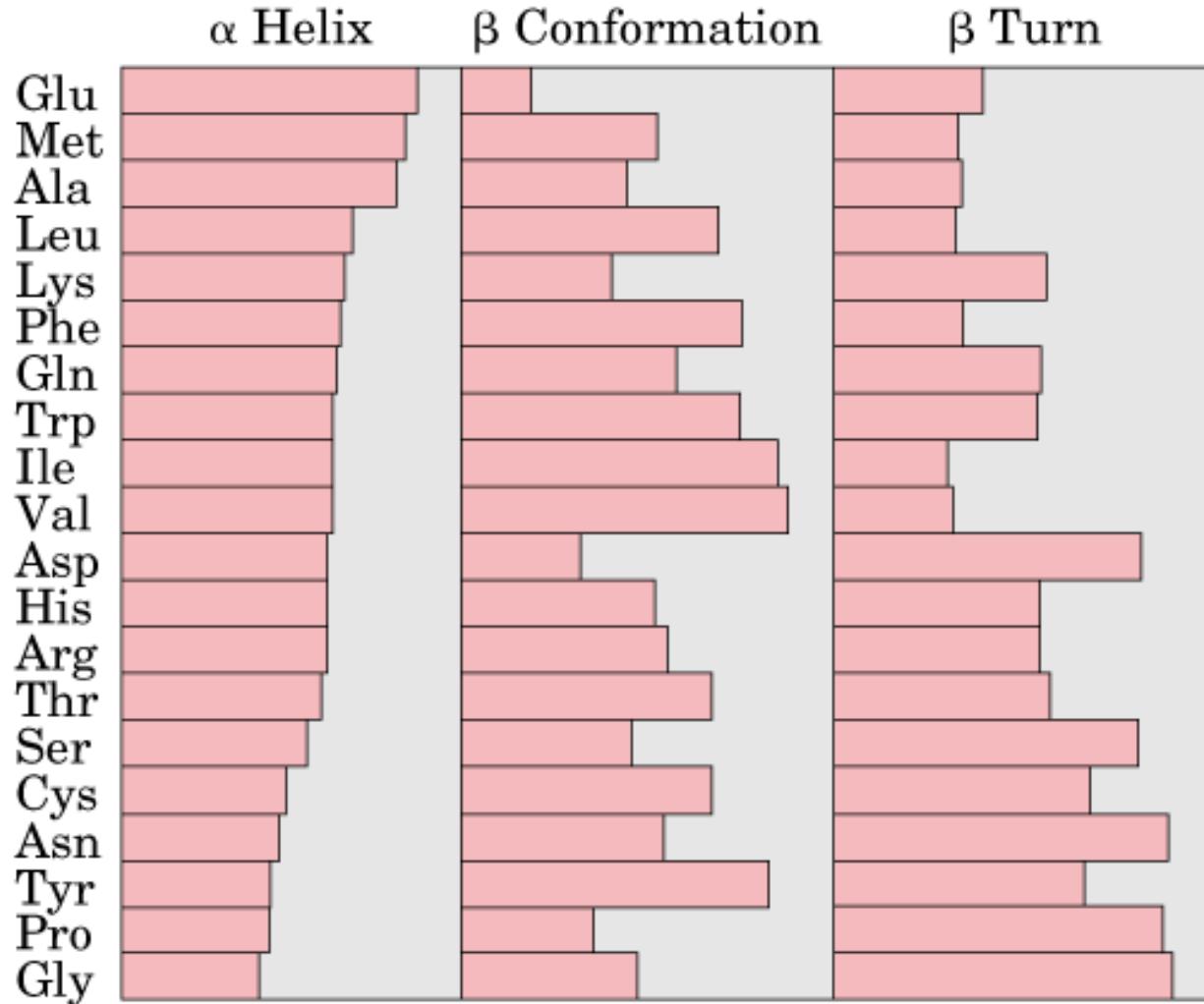
Estructura Cuaternaria

La estructura cuaternaria corresponde a la **unión de distintas cadenas polipeptídicas** para formar un complejo proteico, el cual **es mantenido por interacciones no-covalentes**.

No todas las proteínas tienen estructura cuaternaria.



Abundancia de Aminoácidos en Estructuras Secundarias



Abundancia de Estructuras Secundarias en Proteínas

table 6–2

Approximate Amounts of α Helix and β Conformation in Some Single-Chain Proteins*

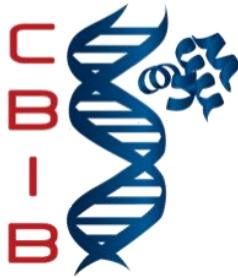
Protein (total residues)	Residues (%)	
	α Helix	β Conformation
Chymotrypsin (247)	14	45
Ribonuclease (124)	26	35
Carboxypeptidase (307)	38	17
Cytochrome c (104)	39	0
Lysozyme (129)	40	12
Myoglobin (153)	78	0

Source: Data from Cantor, C.R. & Schimmel, P.R. (1980) *Biophysical Chemistry*, Part I: *The Conformation of Biological Macromolecules*, p. 100, W.H. Freeman and Company, New York.

*Portions of the polypeptide chains that are not accounted for by α helix or β conformation consist of bends and irregularly coiled or extended stretches. Segments of α helix and β conformation sometimes deviate slightly from their normal dimensions and geometry.

PARTE 2:

Genómica en Cáncer



Enzyme Informatics and Phylogenomics Lab



\$\$\$



FONDECYT
Fondo Nacional de Desarrollo
Científico y Tecnológico



Thomas Peterson
Maricel Kann



Manuel Rozas



Kathleen Whitlock
Christian Cortes
Ramon Latorre



Eduardo Agosin
Sebastián Rubat
Arturo Borzutzky



University of California
San Francisco

Patricia Babbitt
Eyal Akiva
Gemma Holliday



**UNIVERSIDAD
ANDRÉS BELLO**

Marcela Giacometto
Luis Velásquez
Claudio Meneses
Andrea Moreno

Definitions

Enzyme Informatics

“Efforts to understand, predict, control and even redesign the function of enzymes ...”

“... allow[s] associations to be made between sequences, structures, enzyme classification, chemical reactions, mechanisms, cellular functions, protein-ligand interactions, protein expression, metabolism, disease, and pharmacology.

Enzyme Informatics. Alderson RG et al. *Curr Top Med Chem.* 2012,
12:1911-19223.

Definitions

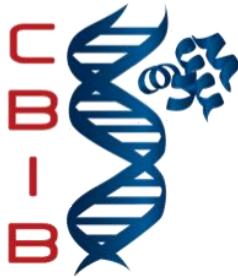
Phylogenomics

- (1) using genomic data to infer phylogenetic relationships and gain insights into the mechanisms of molecular evolution
- (2) using multi-species comparisons and phylogenetics to infer putative functions for DNA or protein sequences.

Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Eisen JA.

Genome Res 1998, 8:163-167.

Overview of the First Phylogenomics Conference. Philippe H and Blanchette M. *BMC Evolutionary Biology* 2007, 7:S1.



Enzyme Informatics and Phylogenomics Lab



Molecular markers in pathologies

- Analysis of mutations in exomes from tumors
- Identification of markers in obese Chilean women
- TIC4GEN – Development of a platform for assisted genetic diagnosis for Latin American population

Discovery and re-engineering of enzymes for biotechnological applications

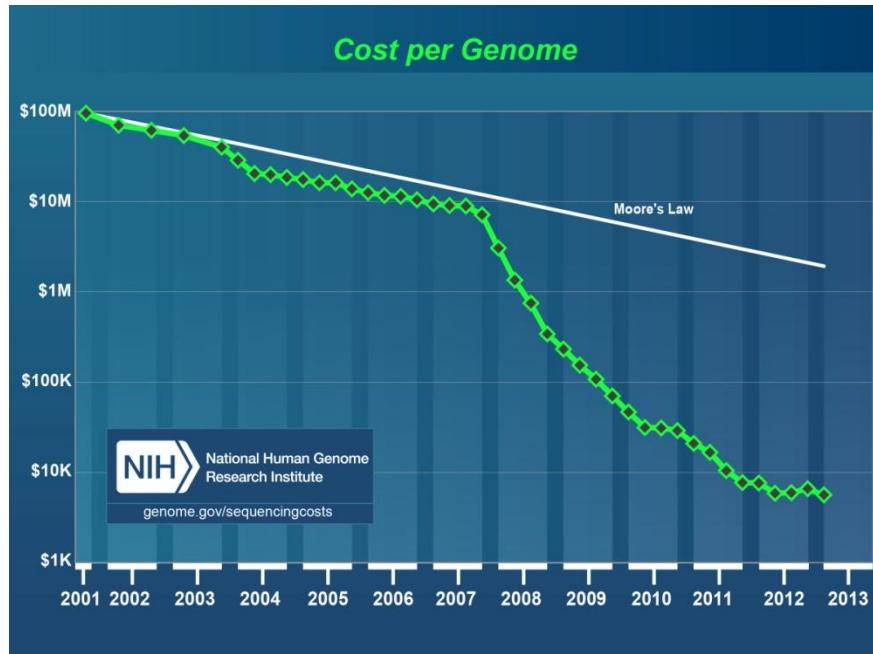
- Synthesis of aroma and flavor molecules in yeast
- Aminotec – Synthesis of amino acids using novel Antarctic bacteria
- Kura Biotec – Novel enzymatic reagent for the analysis of drugs and steroids

Evolutionary studies of protein superfamilies

TRP, GnRH, Omp, Ribokinase, among others

Sequencing Costs Keep Going Down!

With the emergence of ever cheaper technologies for DNA sequencing, today there are thousands of tumor samples that have been subjected to whole-genome or whole-exome sequencing.



April 2004: US\$20,442,576
Q2 2014: US\$1,000

20,000 times
decrease in 10
years!!!

Sequencing Technologies

Now



HiSeq X TEN
Human Genome per
\$1,000

Future : US\$?



At UNAB



MiSeq
Human Genome at
40X per ~\$12,000



Claudio Meneses



Oxford Nanopore
Cost of Machine: US\$1,000
Human Genome per ???



Andrea Moreno Switt

Mutation Data Used in Our Study

Whole-exome somatic mutations for 4327 patients affected by 19 different types of Cancer were obtained from **The Cancer Genome Atlas** initiative of the National Cancer Institute of the USA.

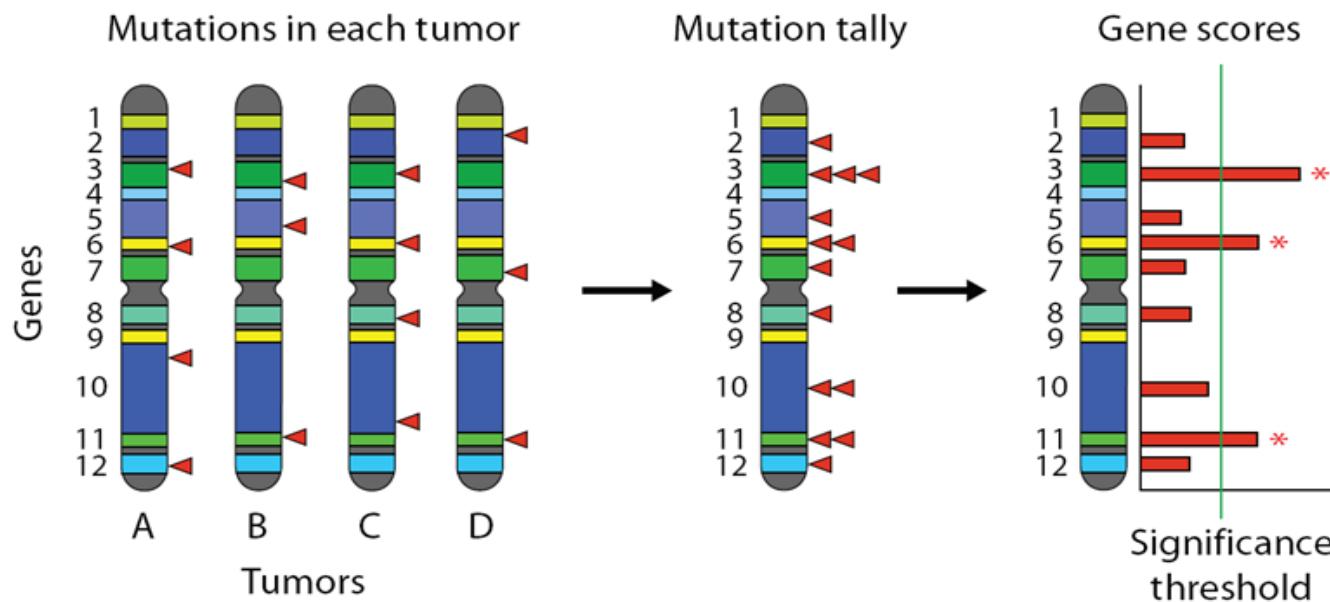
Collins FS &
Barker AD. *Sci Am* 2007, 296:50

Type of Cancer	Abbrev.	No of Patients
Bladder Urothelial Carcinoma	BCLA	28
Cervical Squamous Cell Carcinoma	CESC	39
Kidney Renal Clear Cell Carcinoma	KIRC	297
Kidney Renal Papillary Cell Carcinoma	KIRP	114
Brain Lower Grade Glioma	LGG	217
Lung Squamous Cell Carcinoma	LUSC	178
Skin Cutaneous Melanoma	SKCM	228
Stomach Adenocarcinoma	STAD	221
Thyroid Adenocarcinoma	THCA	323
Prostate Adenocarcinoma	PRAD	83
Ovarian Serous Cystadenocarcinoma	OV	316
Breast Invasive Carcinoma	BRCA	772
Colon Adenocarcinoma	COAD	155
Glioblastoma Multiforme	GBM	284
Lung Adenocarcinoma	LUAD	248
Rectum Adenocarcinoma	READ	73
Uterine Corpus Endometrioid Carcinoma	UCEC	248
Head and Neck Squamous Cell Carcinoma	HNSC	306
Acute Myeloid Leukemia	LAML	197

Selection of Significantly Mutated Genes

We normalized the data taking into account mutational heterogeneity of the samples by applying the MutSigCV method.

It uses patient-specific mutation frequency, and gene-specific background mutation rates to determine statistically the genes that are significantly mutated in each sample.

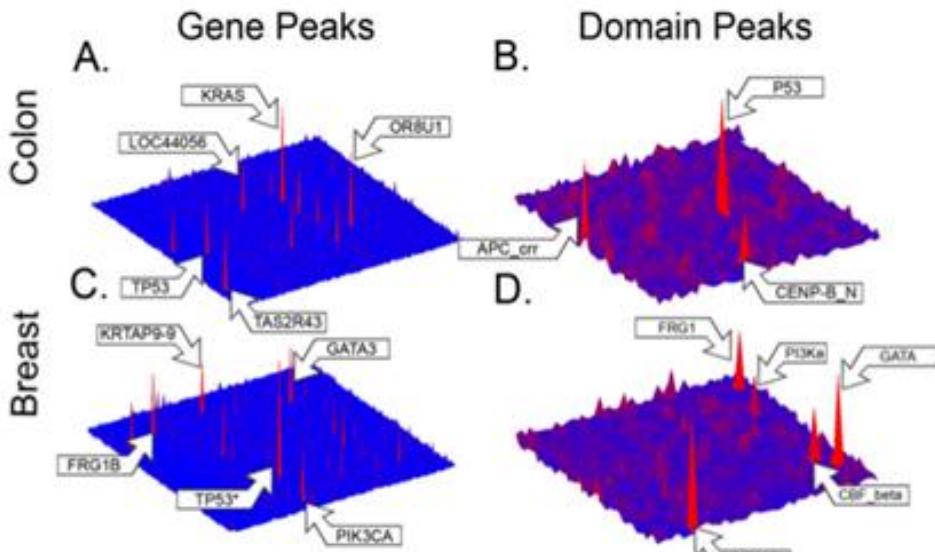


Lawrence, M. et al. *Nature* 2013, 499:214

Significantly Mutated Genes

Type of Cancer	Abbreviation	No of Patients	Q < 0.01	Significant mutations
Bladder Urothelial Carcinoma	BCLA	28	0	0
Cervical Squamous Cell Carcinoma	CESC	39	0	0
Kidney Renal Clear Cell Carcinoma	KIRC	297	6	370
Kidney Renal Papillary Cell Carcinoma	KIRP	114	2	11
Brain Lower Grade Glioma	LGG	217	12	603
Lung Squamous Cell Carcinoma	LUSC	178	12	523
Skin Cutaneous Melanoma	SKCM	228	46	2010
Stomach Adenocarcinoma	STAD	221	22	607
Thyroid Adenocarcinoma	THCA	323	19	311
Prostate Adenocarcinoma	PRAD	83	0	0
Ovarian Serous Cystadenocarcinoma	OV	316	2	282
Breast Invasive Carcinoma	BRCA	772	67	1879
Colon Adenocarcinoma	COAD	155	12	526
Glioblastoma Multiforme	GBM	284	11	458
Lung Adenocarcinoma	LUAD	248	168	6642
Rectum Adenocarcinoma	READ	73	3	181
Uterine Corpus Endometrioid Carcinoma	UCEC	248	26	181
Head and Neck Squamous Cell Carcinoma	HNSC	306	34	1155
Acute Myeloid Leukemia	LAML	197	0	0
TOTAL		4327	442	15739

A Very Brief Intro about Studying Genes in Isolation vs Studying Them in Families

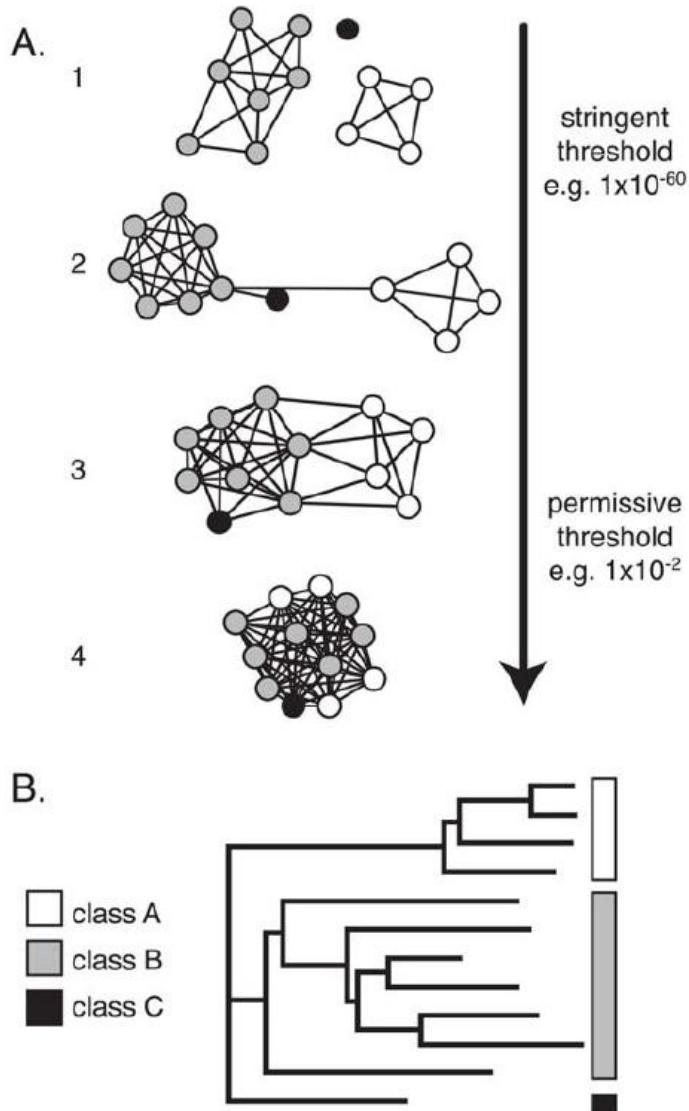


Nehrt et al., *BMC Genomics* 2012, 13:S9

If mutations in genes containing homologous domains are studied together :

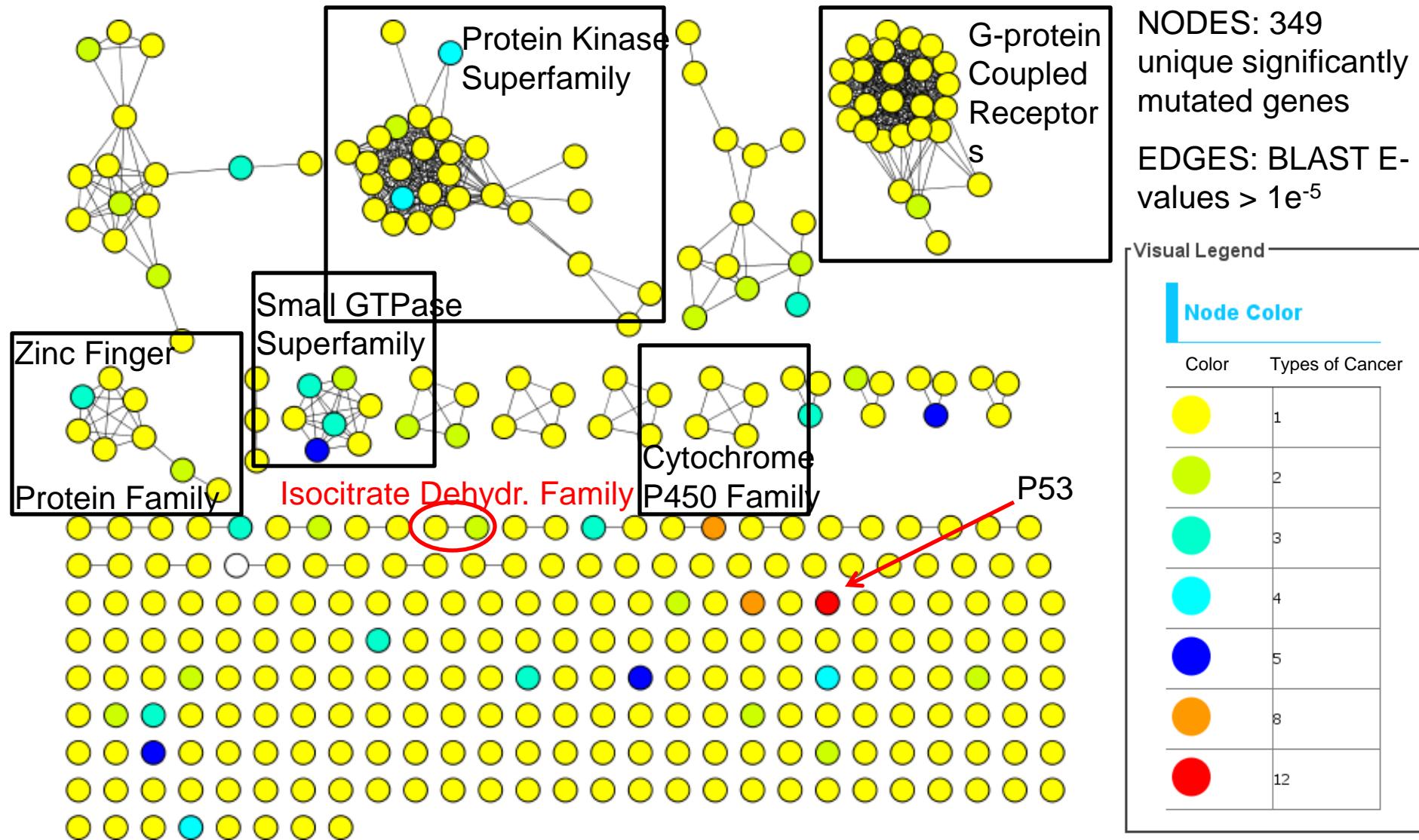
- (1) they can reveal molecular changes that under a gene-centric approach would have been too rare to be of interest, or else
- (2) reveal mutations that at the gene level appear to be of interest, but at the superfamily level are shown to reside in regions of poor conservation.

A Very Brief Intro about Sequence Similarity Networks

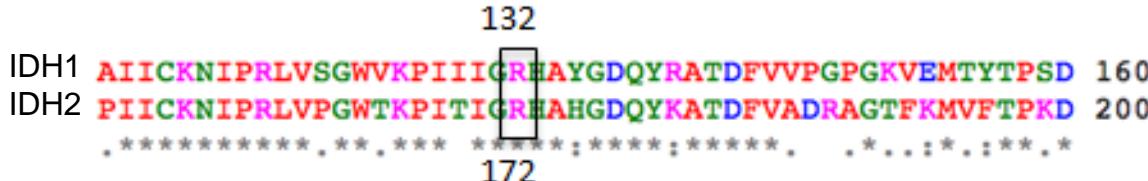


- Fast to compute (only need all-against-all BLAST calculations)
- As with other biological networks, it is robust to missing data
- Correlates well with phylogenetic trees
- Allows easy identification of clades
- Ideal for creating visual summaries of families/superfamilies of proteins

Sequence Similarity Network of Significantly Mutated Genes in Cancer



Significant Mutations in the Isocitrate Dehydrogenase Family



IDH1 mutants

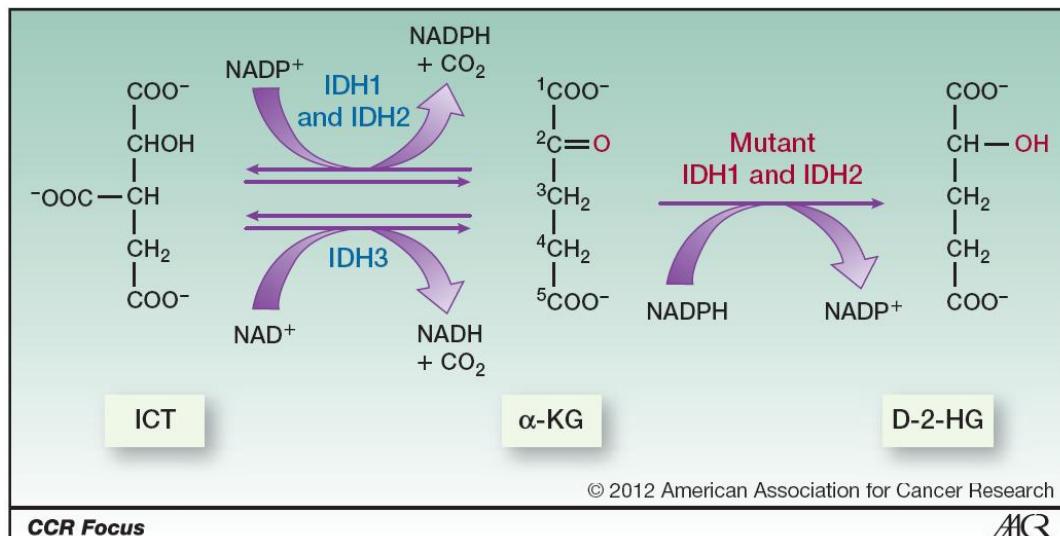
14 p.R132C
6 p.R132G
162 p.R132H
2 p.R132L
5 p.R132S

IDH2 mutants

6 p.R172K
1 p.R172M
1 p.R172S
1 p.R172W

Types of Cancer involved:

GBM, LGG and LAML.



D-2-HG competes with α-KG as redox cofactor in dioxygenases, including epigenetic regulators such as the Jumonji family of histone demethylases and Tet2.

This results in hypermethylation of histones and DNA and thus alterations in transcription.

Differential Binding of NADPH and keto-glutarate in Isocitrate Dehydrogenase

